Perceptual Foundations of Abstract Thought

by

Daniel J. Casasanto

G.P.D., Vocal Performance
Peabody Conservatory of the Johns Hopkins University, 1996

B.A., English and Vocal Performance
Oberlin College and Oberlin Conservatory of Music, 1993

Submitted to the Department of Brain and Cognitive Sciences in Partial Fulfillment
of the Requirements for the Degree of

Doctor of Philosophy In Cognitive Science

at the

Massachusetts Institute of Technology

September, 2005

Signature of Author:

____

Department of Brain and Cognitive Sciences
7 September, 2005

Certified by:

____

Lera Boroditsky
Assistant Professor of Psychology, Stanford University
Thesis Supervisor

Accepted by:

____

Earl K. Miller
Picower Professor of Neuroscience
Chairman, Department Graduate Committee

Abstract

How do people think about things they can never see or touch? The ability to invent and reason about domains such as time, ideas, or mathematics is uniquely human, and is arguably the hallmark of human sophistication. Yet, how people mentally represent these abstract domains has remained one of the great mysteries of the mind. This dissertation explores a potential solution: perhaps the mind recruits old structures for new uses. Perhaps sensory and motor representations that result from physical interactions with the world (e.g., representations of physical space) are recycled to support our thinking about abstract phenomena. This hypothesis is motivated, in part, by patterns observed in language: in order to talk about abstract things, speakers often recruit metaphors from more concrete or perceptually rich domains. For example, English speakers often talk about time using spatial language (e.g., a *long* vacation; a *short* meeting). Cognitive linguists have argued such expressions reveal that people conceptualize abstract domains like time metaphorically, in terms of space. Although linguistic evidence for this Conceptual Metaphor Theory is abundant, the necessary nonlinguistic evidence has been elusive.

In two series of experiments, I investigated whether mental representations that result from physical experience underlie people's more abstract mental representations, using the domains of space and time as a testbed. New experimental tools were developed in order to evaluate Conceptual Metaphor Theory as an account of the evolution and structure of abstract concepts, and to explore relations between language and nonlinguistic thought. Hypotheses about the way people represent space and time were based on patterns in metaphorical language, but were tested using simple psychophysical tasks with nonlinguistic stimuli and responses. Results of the first set of experiments showed that English speakers incorporate irrelevant spatial information into their estimates of time (but

not vice versa), suggesting that people not only talk about time using spatial language, but also think about time using spatial representations. The second set of experiments showed that (a) speakers of different languages rely on different spatial metaphors for duration, (b) the dominant metaphor in participants' first languages strongly predicts their performance on nonlinguistic time estimation tasks, and (c) training participants to use new spatiotemporal metaphors in language changes the way they estimate time. Together, these results demonstrate that the metaphorical language people use to describe abstract phenomena provides a window on their underlying mental representations, and also shapes those representations. The structure of abstract domains such as time appears to depend, in part, on both linguistic experience and on physical experience in perception and motor action.

Acknowledgments

I am deeply grateful for the almost unimaginable good fortune to have spent my graduate years as part of the cognitive science communities at MIT, Harvard, and Stanford, and to the people who make these communities intellectually vibrant. In particular, I thank the members of my thesis committee, Susan Carey, Molly Potter, and Josh Tenenbaum, not only for supporting me during the birthing of this dissertation, but also for challenging and guiding me over the past years. I am especially indebted to my thesis supervisor, Lera Boroditsky, who is a continual source of inspiration. To her credit (and my amazement), Lera allowed me to do experiments in her lab that were designed to show how crazy some of her ideas were -- and she never once said, "I told you so" when we saw the results.

I also owe a special debt to Steven Pinker, whose ideas resonate throughout these pages, and whose books convinced me to abandon a perfectly respectable career path to pursue cognitive science. (This is something I never confessed to Steve, for fear he would think I was a just groupie, and not a serious student. Now that I've completed 5 years of grad school, I trust he'll believe that I'm a sincere student of the mind -- as well as a groupie.) I thank Lila Gleitman for believing that an opera singer could contribute to the study of language and cognition, and for encouraging me to try. I'm grateful to Herb Clark for being incredibly generous with his time, and for sharing his remarkable mind -- which is even sharper than his tongue.

If academic siblings are as important for our development as academic parents, I am twice blessed. In particular, I thank David Badre, Jodi Davenport, Dan Grodner, Webb Phillips, Jaemin Rhee, Doug Rhode, Richard Russell, Ned Sahin, Rebecca Saxe, Jon Winawer, and Nathan Witthoft at MIT, as well as members of LDS at Harvard, and members of SLUGs and the Tversky lab at Stanford for their willingness to argue about ideas, hash out experimental designs, puzzle about data, and to make work feel like play.

One of the best parts of my graduate experience has been working with undergraduate RAs, who have astonished me with their ingenuity, curiosity, and dedication. I am grateful to all of the undergrads who contributed to this dissertation research, especially Jorge Bravo, Shima Goswami, Jesse Greene, Shirley Li, Steve Malliaris, Genevieve Ricart, Lynne Salameh, and Alice Suh at MIT, and Wes Mateo, Morgan Springer, and Itamar Rosenn at Stanford.

The cross-linguistic studies reported here would not have been possible without the contributions of native Greek, Indonesian, and Spanish speakers who generously shared their time and expertise, including Olga Fotokopoulu and Ria Pita at the Aristotle University of Thessaloniki, Uri Tadmoor at the Max Planck Center for Evolutionary Anthropology in Jakarta, Callie Piedade Shona Watkins at Harvard, and Ilia Santiago-Diaz, Wendy Ham, and Simon Bocanegra-Thiel at MIT.

On the business end of things, I am grateful for the support of an NSF graduate research fellowship and an NSF dissertation improvement grant. I thank the Vivian Smith Advanced Studies Institute for the opportunity to conduct research in Greece, and the Exchange Scholars program for allowing me to spend the past year at Stanford. Closer to home, I thank Denise Heintze for always looking out for me, and for tolerating my somewhat Italian disposition toward schedules and deadlines (surely evidence of cultural relativity).

Finally, I thank my entire family for their unwavering love and support (no matter how many times I went to graduate school) especially my mother, Joyce Williams, and my grandparents, John and Anna Williams. I dedicate this dissertation to the memory of my father, Peter Casasanto, and my grandparents, Daniel and Anna Casasanto.

Contents

Chapter 1: Introduction

1.1 Time as an abstract domain

> *For what is time? Who can readily and briefly explain this? Who can even in thought comprehend it, so as to utter a word about it? (…) If no one asks me, I know: if I wish to explain it to one who asketh, I know not.*
>
> Saint Augustine, *Confessions*, Book 11

How long will it take you to read this dissertation? The objective time, as measured by the clock, might depend on whether you're scrutinizing every detail, or just skimming to get the main ideas. The subjective time might vary according to physiological factors like your pulse and body temperature (Cohen, 1967; Ornstein, 1969), psychological factors like how much the text engages your interest and attention (Glicksohn, 2001; James, 1890; Zakay & Block, 1997), and some surprising environmental factors such as the size of the room you're sitting in (DeLong, 1981).

Although subjective duration is among the earliest topics investigated by experimental psychologists (Mach, 1886), the cognitive sciences have yet to produce a comprehensive theory of how people track the passage of time, or even to agree on a set of principles that consistently govern people's duration estimates. A passage from a review by Zakay and Block (1997) illustrates the current state of confusion:

> People may estimate filled durations as being longer than empty durations, but sometimes the reverse is found. Duration judgments tend to be shorter if a more difficult task is performed than if an easier task is performed, but again the opposite has also been reported. People usually make longer duration estimates for complex than for simple stimuli, although some researchers have found the opposite. (pg. 12)

What makes time perception so difficult to understand? Ornstein (1969) suggests the very idea that time can be *perceived* through the senses is misleading:

> One major reason for the continuing scattering of [researchers'] effort has been that time is treated as if it were a sensory process. If time were a sensory process like vision…we would have an 'organ' of time experience such as the eye. (pg. 34)

Although time is not something we can see or touch, we often talk about it as if it were (Boroditsky, 2000; Clark, 1973; Gruber, 1965; Jackendoff, 1983; Lakoff & Johnson, 1980). Consider the following pair of sentences:

i. They moved the truck forward two meters.
ii. They moved the meeting forward two hours.

The truck in sentence i. is a physical object which can move forward through space, and whose motion we might see, hear, or feel, from the staring point to the ending point. By contrast, there is no literal motion described in sentence ii. The meeting is not translated through space, and there is no way to experience its 'movement' through time via the senses. Events that occur in time are more abstract than objects that exist in space insomuch as we typically have richer perceptual evidence for the spatial than for the temporal[1].

In the chapters that follow, I will show that (a) the language people typically use to talk about duration reveals important links between the abstract domain of time and the relatively concrete domain of space, (b) temporal representation must be understood, in part, in terms of spatial representations, and (c) the domains of space and time provide a testbed for hypotheses about the evolution and structure of abstract concepts.

1.2 Metaphor and the problem of abstract thought

The mystery of how people come to mentally represent abstract domains such as time, ideas, or mathematics has engaged scholars for centuries, sometimes leading to proposals that seem unscientific by modern standards. Plato (*Meno*, ca. 380 B.C.E.) argued that we cannot acquire abstract concepts like *virtue* through instruction, and since babies are not born knowing them, it must be that we recover such concepts from previous incarnations of

[1] Some of our spatial representations may be quite abstract, as well. For example, our conception of the Milky Way galaxy's breadth is no more grounded in direct experience than our conception of its age.

our souls. Charles Darwin contended that evolution can explain the emergence of abstract thought without recourse to reincarnation, yet it is not immediately obvious how mental capacities that would have been superfluous for our Pleistocene forebears could have been selected for. What selection pressures could have resulted in our ability to compose symphonies, invent calculus, or imagine time travel? How did foragers become physicists in an eyeblink of evolutionary time? The human capacity for abstract thought seems to far exceed what could have benefited our predecessors, yet natural selection can only effect changes that are immediately useful. The apparent evolutionary uselessness of human intelligence drove Alfred Wallace, Darwin's co-founder of the theory of evolution by natural selection, to abandon their theory and invoke creationism to explain our capacity for abstract thought (Darwin, 1859/1998, 1874/1998; Gould, 1980; Pinker, 1997; Wallace, 1870/2003)[2].

Darwin's own formulation of evolutionary theory points toward an elegant potential solution to Wallace's dilemma: sometimes organisms recycle old structures for new uses. An organ built via selection for a specific role may be fortuitously suited to perform other unselected roles, as well. For example, the fossil record suggests that feathers were not originally 'designed' for flying. Rather, they evolved to regulate body temperature in small running dinosaurs, and were only later co-opted for flight (Gould, 1991). The process of adapting existing structures for new functions, which Darwin (1859/1993) gave the misleading name *preadaptation*, was later dubbed *exaptation* by evolutionary biologist Steven Jay Gould and paleontologist Elisabeth Vrba (1982). Gould argued that this process may explain the origin of many improbable biological and psychological structures.

[2] Cultural evolution alone cannot explain our capacity for abstract thought because, as Wallace noted, members of "stone age" societies who were given European educations manifested abilities to similar those of modern Europeans: the latent capacity to read, to perform Western art music, etc. was present in the minds of people whose cultures had never developed these abstract forms of expression.

Can exaptation account for mental abilities in humans that could not have been selected for directly? If so, how might this have happened? Which *ad*apted capacities might abstract domains be *ex*apted from? Steven Pinker (1997) sketched the following proposal:

> Suppose ancestral circuits for reasoning about space and force were copied, the copies' connections to the eyes and muscles were severed, and references to the physical world were bleached out. The circuits could serve as a scaffolding whose slots are filled with symbols for more abstract concerns like states, possessions, ideas, and desires. (pg. 355)

As evidence that abstract domains arose from circuits designed for reasoning about the physical world, Pinker appeals to patterns observed in language. Many linguists have noted that when people talk about states, possessions, ideas, and desires, they do so by co-opting the language of intuitive physics (Clark, 1973, Gibbs, 1994; Gruber, 1965; Jackendoff, 1983; Lakoff & Johnson, 1980; Langacker, 1987; Talmy, 1988). In particular, words borrowed from physical domains of space, force, and motion, give rise to metaphors for countless abstract ideas[3].

1*l* a *high* shelf
1*m* a *high* price

2*l* a *big* building
2*m* a *big* debate

3*l* *forcing* the door
3*m* *forcing* the issue

4*l* *pushing* the button
4*m* *pushing* the limit

5*l* *keeping* the roof *up*
5*m* *keeping* appearances *up*

---

[3] The concrete objects described in the literal sentences (e.g., shelf, building, door, button, roof) belong to a different ontological category than the abstract entities in the metaphorical examples, according a test of what physical relations they can sensibly be said to enter into. For example, it is sensible to say "the cat sat on the shelf / building / door / button / roof", but it may not be sensible to say that "the cat sat on the price / debate / issue / limit / appearance." This test is similar to a test of sensible predicates for concrete vs. abstract entities the devised by Fred Sommer (1963).

For each pair above, sentence *l* illustrates a literal use and sentence *m* a metaphorical use of the italicized words. Based on such patterns, linguists have argued that people create abstract domains by importing structure from concepts grounded in physical experience. Although anticipated by others (e.g., Lafargue, 1898/1906), this idea appears to have been first articulated as the Thematic Relations Hypothesis (TRH) in 1965, by Jeffery Gruber in his MIT dissertation. TRH was later elaborated by Jackendoff (1972; 1983) who wrote:

> The psychological claim behind [Gruber's linguistic discovery] is that the mind does not manufacture abstract concepts out of thin air…it adapts machinery that is already there, both in the development of the individual organism and in the evolutionary development of the species. (1983, pg. 188-9)

Not all theorists agree on the significance of metaphorical language for theories of mental representation. Gregory Murphy (1996; 1997) raised concerns about both the vagueness of the psychological processes suggested by linguists and about the limitations of purely linguistic evidence for metaphoric conceptual structure. Murphy (1996) proposed that linguistic metaphors may merely reveal structural similarities between mental domains: not causal relationships. He argued that in the absence of corroborating nonlinguistic evidence, his *Structural Similarity* proposal should be preferred on grounds of simplicity: his view posits that all concepts are represented independently, on their own terms, whereas the metaphoric alternative posits complex concepts that are structured interdependently. It is evident that people *talk* about abstract domains in terms of relatively concrete domains, but do they really *think* about them that way?

1.3 Current evidence for Conceptual Metaphor

The claim that conventionalized metaphors in language reveal the structure of abstract concepts is often associated with linguist George Lakoff and philosopher Mark Johnson (L&J). According to L&J, Conceptual Metaphor theory is one of "three major findings of cognitive science" (1999, pg. 3). However, this claim is supported almost entirely by linguistic evidence, which is supplemented by one computational model providing an *in principle* demonstration that the meanings of some linguistic metaphors can be represented by the same neural network that models physical motion (Narayanan, 1997). Despite the impressive body of linguistic theory and data they summarize, L&J offer little evidence that the importance of metaphor extends beyond language. Pinker (1997) was more realistic in his evaluation of the available evidence when he called the idea that abstract thought is an exaptation from physical domains “just an avowal of faith” among scientists who believe that the mind must ultimately be explicable as a product of natural selection (pg. 301).

Boroditsky (2000) conducted some of the first behavioral tests of Conceptual Metaphor Theory. Her tasks capitalized on the fact that in order to talk about spatial or temporal sequences, speakers must adopt a particular frame of reference. Sometimes we use expressions that suggest we are moving through space or time (e.g., *we’re approaching Maple Street*; *we’re approaching Christmas*). Alternatively, we can use expressions that suggest objects or events are moving with respect to one another (*Maple Street comes before Elm Street*; *Christmas comes before New Year’s*). In one experiment, Boroditsky found that priming participants to adopt a given spatial frame of reference facilitated their interpretation of sentences that used the analogous temporal frame of reference. Importantly, the converse was not found: temporal primes did not facilitate interpreting spatial sentences. This priming asymmetry parallels a well established asymmetry in linguistic metaphors: people talk about the abstract in terms of the concrete (e.g., time in terms of space) more than the other way around (Lakoff & Johnson, 1980). Based on these results Borodisky proposed a refinement of Conceptual Metaphor Theory, the Metaphoric Structuring view, according to which (a) the domains of space and time share conceptual structure, and (b) spatial information is useful (though not necessary) for thinking about time. A second set of experiments showed that real-world spatial situations (e.g., standing at the beginning, middle, or end of a cafeteria line) and even imaginary spatial scenarios can influence how people interpret spatiotemporal metaphors (Boroditsky & Ramscar, 2002). These studies rule out what Boroditsky (2000) calls the Dubious View, that space-time metaphors are simply "etymological relics with no psychological consequences" (pg. 6).

If people use spatial schemas to think about time, as suggested by metaphors in language, then do people who use different kinds of spatiotemporal metaphors in their native tongues think about time differently? To find out, Boroditsky (2001) compared performance on space-time priming tasks in speakers of English, a language which typically describes time as horizontal, and speakers of Mandarin Chinese, which also commonly uses vertical spatiotemporal metaphors. English speakers were faster to judge sentences about temporal succession (e.g., *March comes earlier than April*) when primed with a horizontal spatial event, but Mandarin speakers were faster to judge the same sentences when primed with a vertical spatial stimulus. This was true despite the fact that all of the sentences were presented in English. In a follow-up study, Boroditsky (2001) trained English speakers to use vertical metaphors for temporal succession (e.g., March is *above* April). After training, their priming results resembled those of the native Mandarin speakers.

Together, Boroditsky's studies provide some of the first evidence that (a) people not only talk about time in terms of space, they also think about it that way, (b) people who use different spatiotemporal metaphors also think about time differently, and (c) learning new spatial metaphors can change the way you mentally represent time. Yet, these conclusions are subject to a skeptical interpretation. Boroditsky's participants made judgments about sentences containing spatial or temporal language. Perhaps their judgments showed relations between spatial and temporal thinking that were consistent with linguistic metaphors only because they were required to process space or time *in language*. Would the same relations between representations of space and time be found if participants were tested on nonlinguistic tasks?

The fact that people communicate via language replete with anaphora, ambiguity, metonymy, sarcasm, and deixis seems irrefutable proof that what we say provides only a thumbnail sketch of what we think. Most theorists posit at least some independence between semantic representations and underlying conceptual representations (Jackendoff, 1972; Katz & Fodor, 1963; Levelt, 1989; c.f., Fodor, 1975). Even those who posit a single, shared 'level' of representation for linguistic meaning and nonlinguistic concepts allow that semantic structures must constitute only a subset of conceptual structures (Chomsky, 1975; Jackendoff, 1983). Because we may think differently when we're using language and when we're not, well-founded doubts persist about how deeply patterns in language truly reflect – and shape – our nonlinguistic thought. According to linguist Dan Slobin (1996):

> Any utterance is a selective schematization of a concept – a schematization that is in some ways dependent on the grammaticized meanings of the speaker's particular language, recruited for the purposes of verbal expression. (pg. 75-76)

Slobin argues that when people are "thinking for speaking" (and presumably for reading or listening to speech), their thoughts are structured, in part, according to their language and its peculiarities. Consequently, speakers of different languages may think differently when they are using language. But how about when people are not thinking for speaking? Eve Clark (2003) asserts that:

> [When people are] thinking for remembering, thinking for categorizing, or one of the many other tasks in which we may call on the representations we have of objects or events – then their representations may well include a lot of material not customarily encoded in their language. It seems plausible to assume that such conceptual representations are nearer to being universal than the representations we draw on for speaking. (pg. 21)

Clark predicts that results may differ dramatically between tests of language-thought relations that use language and those that do not:

> …we should find that in tasks that require reference to representations in memory that don't make use of any linguistic expression, people who speak different languages will respond in similar, or even identical, ways. That is, representations for nonlinguistic purposes may differ very little across cultures or languages. (2003, pg. 22)

Clark adds:

> Of course, finding the appropriate tasks to check on this without any appeal to language may prove difficult. (2003, pg. 22)

Clark's skepticism echoes concerns raised by Papafrougou, Massey, and Gleitman (2002) regarding the difficulty of studying the language-thought interface:

> …domains within which language might interestingly influence thought are higher-level cognitive representations and processes, for instance, the linguistic encoding of time […] A severe difficulty in investigating how language interfaces thought at these more "significant" and "abstract" levels has been their intractability to assessment. As so often, the deeper and more culturally resonant the cognitive or social function, the harder it is to capture it with the measurement and categorization tools available to psychologists. (pg. 191-192)

For the studies reported here, new experimental tools were developed in order to (a) evaluate Conceptual Metaphor Theory as an account of the structure and evolution of abstract concepts, and (b) to investigate relations between language and nonlinguistic thought, using the domains of space and time as a testbed. Specifically, the goal of Experiments 1-7 was to determine whether English speakers' nonlinguistic mental representations of space and time are related in ways predicted by linguistic metaphors. Experiments 8-10 tested whether nonlinguistic mental representations of time differ among speakers of different languages, in ways consistent with their language-particular metaphors. Experiment 11 investigated whether language can *cause* differences in the nonlinguistic mental representations of time to arise among speakers of different languages.

These experiments used novel psychophysical tasks with nonlinguistic stimuli and responses in order to mediate between two theoretical positions, one which posits *shallow* and the other *deep* relations between language and nonlinguistic thought (Table 1):

Table 1.

The Shallow View:	**The Deep View:**
i. Language reflects the structure of the mental representations that speakers form for the purpose of using language. These are likely to be importantly different, if not distinct, from the representations people use when they are thinking, perceiving, and acting without using language.	**i.** Language reflects the structure of the mental representations that speakers form for the purpose of using language. These are likely to be similar to, if not overlapping with, the representations people use when they are thinking, perceiving, and acting without using language.
ii. Language may influence the structure of mental representations, but only (or primarily) during language use.	**ii.** Patterns of thinking established during language use may influence the structure of the mental representations that people form even when they're not using language.
iii. Cross-linguistic typological differences are likely to produce 'shallow' behavioral differences on tasks that involve language or high-level cognitive abilities (e.g., explicit categorization). However, such behavioral differences should disappear when subjects are tested using nonlinguistic tasks that involve low-level perceptuo-motor abilities.	**iii.** Some cross-linguistic typological differences are likely to produce 'deep' behavioral differences, observable not only during tasks that involve language or high-level cognitive abilities, but also when subjects are tested using nonlinguistic tasks that involve low-level perceptuo-motor abilities.
iv. Although the semantics of languages differ, speakers' underlying conceptual and perceptual representations are, for the most part, universal.	**iv.** Where the semantics of languages differ, speakers' underlying conceptual and perceptual representations may differ correspondingly, such that language communities develop idiosyncratic conceptual repertoires.

Chapter 2: Do people think about time in terms of space?

Do people use mental representations of space in order to mentally represent time, as metaphors in language suggest they do, even when they're not using language? The first six experiments reported here test the hypothesis that temporal thinking depends on spatial thinking, but not vice versa. The seventh experiment tests the specificity of spatial representations that people use to think about time. In each task, participants viewed simple nonlinguistic, non-symbolic stimuli (i.e., lines or dots) on a computer screen, and estimated either their duration or their spatial displacement. Durations and displacements were fully crossed, so there was no correlation between the spatial and temporal components of the stimuli. As such, one stimulus dimension served as a distractor for the other: an irrelevant piece of information that could potentially interfere with task performance. Patterns of cross-dimensional interference were analyzed to reveal relationships between spatial and temporal representations.

Broadly speaking, there are three possible relationships between people's mental representations of space and time. First, the two domains could be *symmetrically dependent*. John Locke (1689/1995) argued that space and time are mutually inextricable in our minds, concluding that, "expansion and duration do mutually embrace and comprehend each other; every part of space being in every part of duration, and every part of duration in every part of expansion" (p. 140). Alternatively, our ideas of space and time could be *independent*. Any apparent relatedness could be due to structural similarities between essentially unrelated domains (Murphy, 1996, 1997). A third possibility is that time and space could be *asymmetrically dependent*. Representations in one domain could be parasitic on representations in the other (Boroditsky, 2000; Gentner, 2001; Gibbs, 1994; Lakoff & Johnson, 1980, 1999).

These three possible relations between space and time predict three distinct patterns of cross-dimensional interference. If spatial and temporal representations are symmetrically dependent on one another, then any cross-dimensional interference should be approximately symmetric: line displacement should modulate estimates of line duration, and vice versa. Alternatively, if spatial and temporal representations are independent, there should be no significant cross-dimensional interference. However, if mental representations of time are asymmetrically dependent on mental representations of space, as suggested by spatiotemporal metaphors in language, then any cross-dimensional interference should be asymmetric: line displacement should affect estimates of line duration more than line duration affects estimates of line displacement.

A total of 125 subjects from the MIT community participated in Experiments 1-7, in exchange for payment. Of these, 35 participants were removed from the analyses reported here for performing the experiment incorrectly (e.g., estimating distance when they were instructed to estimate duration), or for excessively poor performance: for each participant, duration estimates were plotted as a function of actual stimulus duration, and distance estimates were plotted as a function of actual stimulus displacement. Participants were excluded if the slope of their duration or distance estimates was less than 0.5, as such poor performance (e.g., indicating that the 5-second lines lasted less than 2.5 seconds) was believed to result from impatience with the repetitive task, rather than genuine inaccuracy. All participants gave informed consent, and all were native monolingual speakers of English according a language background questionnaire (i.e., English was the only language they learned before age 5, and was their strongest language at time of test).

Studies were approved by the MIT institutional review board's committee on the use of humans as experimental subjects[4].

2.1 Experiment 1: Growing Lines

Materials

Lines of varying lengths were presented on a computer monitor (resolution=1024x768 pixels, dpi=72), for varying durations. Durations ranged from 1000 milliseconds to 5000 milliseconds in 500 millisecond increments. Displacements ranged from 200 to 800 pixels in 75 pixel increments. Nine durations were fully crossed with nine displacements to produce 81 distinct line types. Lines 'grew' horizontally across the screen one pixel at a time, from left to right, along the vertical midline. Lines started growing 112 pixels from the left edge of the monitor on average, but the starting point of each line was jittered with respect to the average starting point (+/- up to 50 pixels), so that the monitor would not provide a reliable spatial frame of reference. Each line remained on the screen until it reached its maximum displacement, and then it disappeared.

Procedure

Participants viewed 162 growing lines, one line at a time. The word "ready" appeared in the center of an otherwise blank screen for two seconds immediately before each line was shown. Immediately after each line was shown, a prompt appeared in either the upper left or lower left corner of the screen indicating that the subject should reproduce either the line's displacement (if an 'X' icon appeared), or its duration (if an 'hourglass' icon appeared). Space trials and time trials were randomly intermixed.

---

[4] A preliminary report on these experiments appeared as Casasanto, D. & Boroditsky, L., (2003). *Do we think about time in terms of space?* Proceedings of the 25th Annual Conference of the Cognitive Sciences Society, Boston, MA, 216-221

To estimate displacement, subjects clicked the mouse once on the center of the X, moved the mouse to the right in a straight line, and clicked the mouse a second time to indicate that they had moved a distance equal to the maximum displacement of the stimulus. Whereas stimuli grew from a jittered starting point on the vertical midline of the screen, responses were initiated at a fixed starting point in either the upper or lower left corner. Thus, the response was translated both vertically and horizontally with respect to the stimulus. To estimate duration, subjects clicked the mouse once on the center of the hourglass icon, waited the appropriate amount of time, and clicked again in the same spot, to indicate the time it took for the stimulus to reach its maximum displacement.

All responses were self-paced. For a given trial, subjects reproduced either the displacement or the duration of the stimulus, never both. Response data were collected for both the trial-relevant and the trial-irrelevant stimulus dimensions, to ensure that subjects were following instructions.

Results and Discussion

Results of Experiment 1 showed that displacement affected estimates of duration, as indicated by a significant correlation between target stimulus displacement and estimated stimulus duration (figure 1a). For stimuli of the same average duration, lines that traveled a shorter distance were judged to take a shorter time, and lines that traveled a longer distance were judged to take a longer time. By contrast, target duration did not affect estimates of spatial displacement (figure 1b). The effect of distance on time estimation was greater than the effect of time on distance estimation, as indicated by a significant difference of correlations (figure 7a). Subjects incorporated irrelevant spatial information in their temporal estimates, but not vice versa. This behavioral asymmetry was predicted based on the asymmetric relationship between time and space in linguistic metaphors.

Overall, estimates of duration and displacement were highly accurate, and about equally accurate in the two domains (figure 1c-d). The asymmetric cross-dimensional interference that was observed cannot be attributed to a difference in the overall accuracy of duration and displacement estimations, as no significant difference in was found.

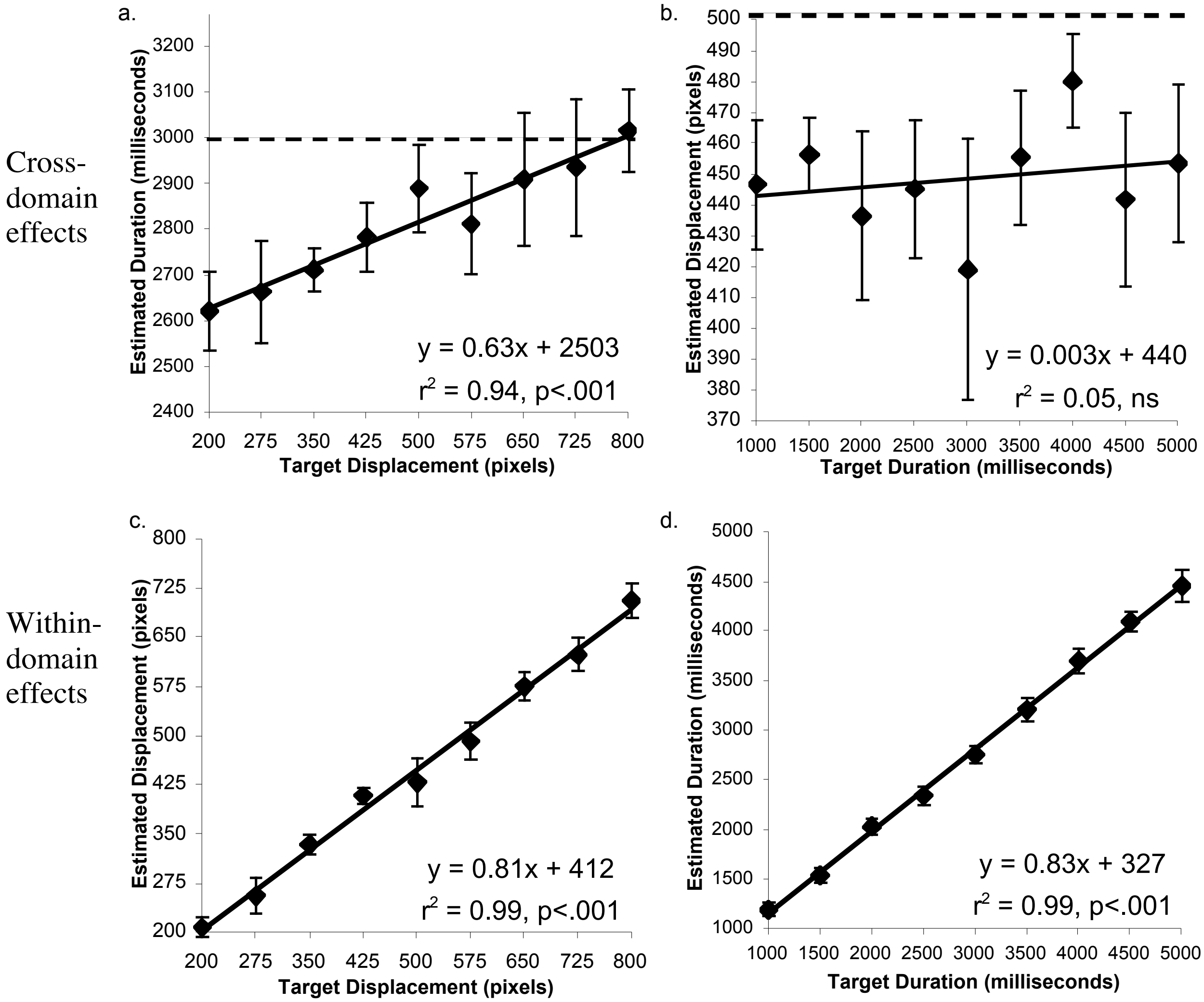

Figure 1. Grand averaged duration and displacement estimates (n=9) for Experiment 1 (Growing Lines). Top: Cross-domain effects. 1a. (left) Effect of displacement on duration estimation. 1b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 1a. and 1b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 1c. (left) Effect of target displacement on estimated displacement. 1d. (right) Effect of target duration on estimated duration.

2.2 Experiment 2: Growing Lines, Selective Attention

What caused this cross-dimensional confusion? In Experiment 1, participants did not know until after each line was presented whether they would need to estimate displacement or duration. They had to attend to both spatial and temporal information, and to update both types of information online throughout the stimulus presentation. If participants were told ahead of time whether they would need to estimate a line's displacement or its duration, would the cross-dimensional interference disappear? Experiment 2 addressed this possibility. Before each line appeared, participants were informed which stimulus dimension they would need to estimate. This gave them the opportunity to attend selectively to the trial-relevant stimulus dimension, and if possible, to ignore the trial-irrelevant dimension.

Materials and Procedure

Stimulus materials were identical to those used in Experiment 1. The procedure was also identical, with one exception. In Experiment 1, the word "ready" appeared for two seconds immediately preceding each line stimulus. In Experiment 2, the word "ready" was replaced either by the word "Space" next to an 'X' icon, or by the word "Time" next to an hourglass icon. These words and symbols indicated whether the subject would need to estimate the displacement or the duration of the next line. Line stimuli, prompts, and responses were exactly as in Experiment 1, thus all stimuli and responses remained entirely nonlinguistic.

Results and Discussion

Results of Experiment 2 replicate those of Experiment 1 (figure 2a-d). Participants were able to disregard line duration when estimating displacement. By contrast, they were unable to ignore line displacement, even when they were encouraged to selectively attend to duration (figure 7b). The cross-dimensional effect of space on time estimation in Experiment 1 was not caused by a task-specific demand for subjects to encode spatial and temporal information simultaneously.

Response data collected for the trial-irrelevant dimension confirmed that participants understood the task, and were not explicitly confusing displacement with duration (i.e., participants were not giving a spatial response when they were supposed to give a temporal response). A blocked version of Experiment 2 was also conducted, in which participants performed all of the distance estimates and then all of the duration estimates (or vice versa). Results of the blocked version did not differ significantly from those of version reported here, in which distance and duration trials were randomly intermixed.

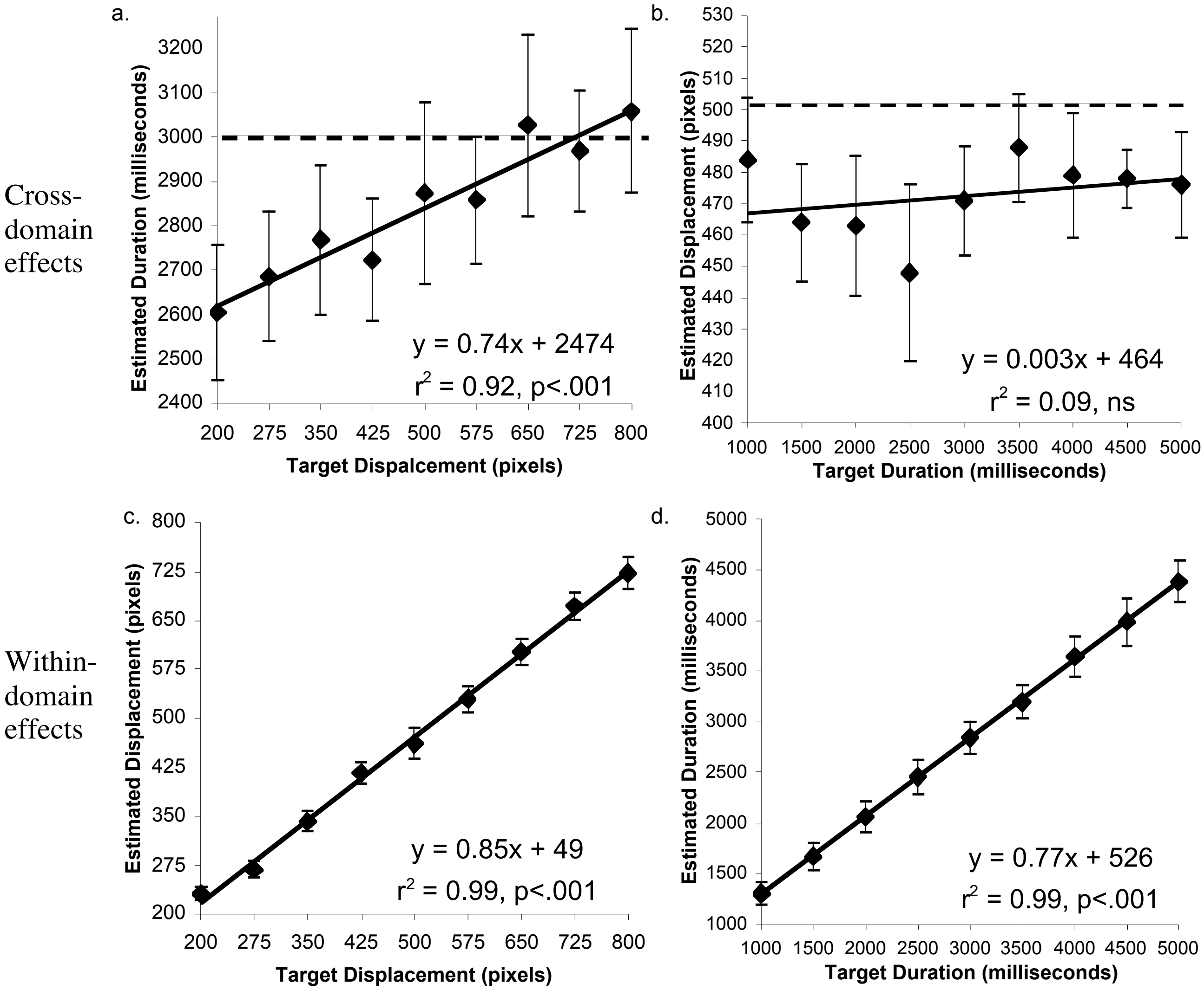

Figure 2. Grand averaged duration and displacement estimates (n=9) for Experiment 2 (Growing Lines, Selective Attention). Top: Cross-domain effects. 2a. (left) Effect of displacement on duration estimation. 2b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 2a. and 2b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 2c. (left) Effect of target displacement on estimated displacement. 2d. (right) Effect of target duration on estimated duration.

2.3 Experiment 3: Growing Lines, Temporal Frame of Reference

Several follow-up experiments were conducted to assess the generality of these results, and to evaluate potential explanations. One concern was that participants may have relied on spatial information to make temporal estimates because stimuli were situated in a constant spatial frame of reference (i.e., the computer monitor). For Experiment 3, stimuli were also situated in a constant temporal frame of reference. Temporal delay periods were introduced preceding and following line presentations, which were proportional to the spatial gaps between the ends of the stimulus lines and the edges of the monitor.

Materials and Procedure

Stimulus materials and procedures were identical to those used in Experiment 2, with the following exception. In the previous experiments, the interval between the disappearance of the 'ready' screen and the appearance of the response prompt varied with stimulus duration. In the present experiment, this interval was fixed at 6400 milliseconds. Stimuli were preceded and followed by a delay period, which was proportional to spatial gap separating the ends of the line stimuli from the left and right edges of the monitor.

Results and Discussion

The same pattern of results was found in Experiment 3 as in the previous experiments (figures 3a-d, figure 7c). Results of Experiment 3 did not differ significantly from those of Experiment 2. The availability of a constant temporal frame of reference did not significantly reduce the influence of distance on time estimation.

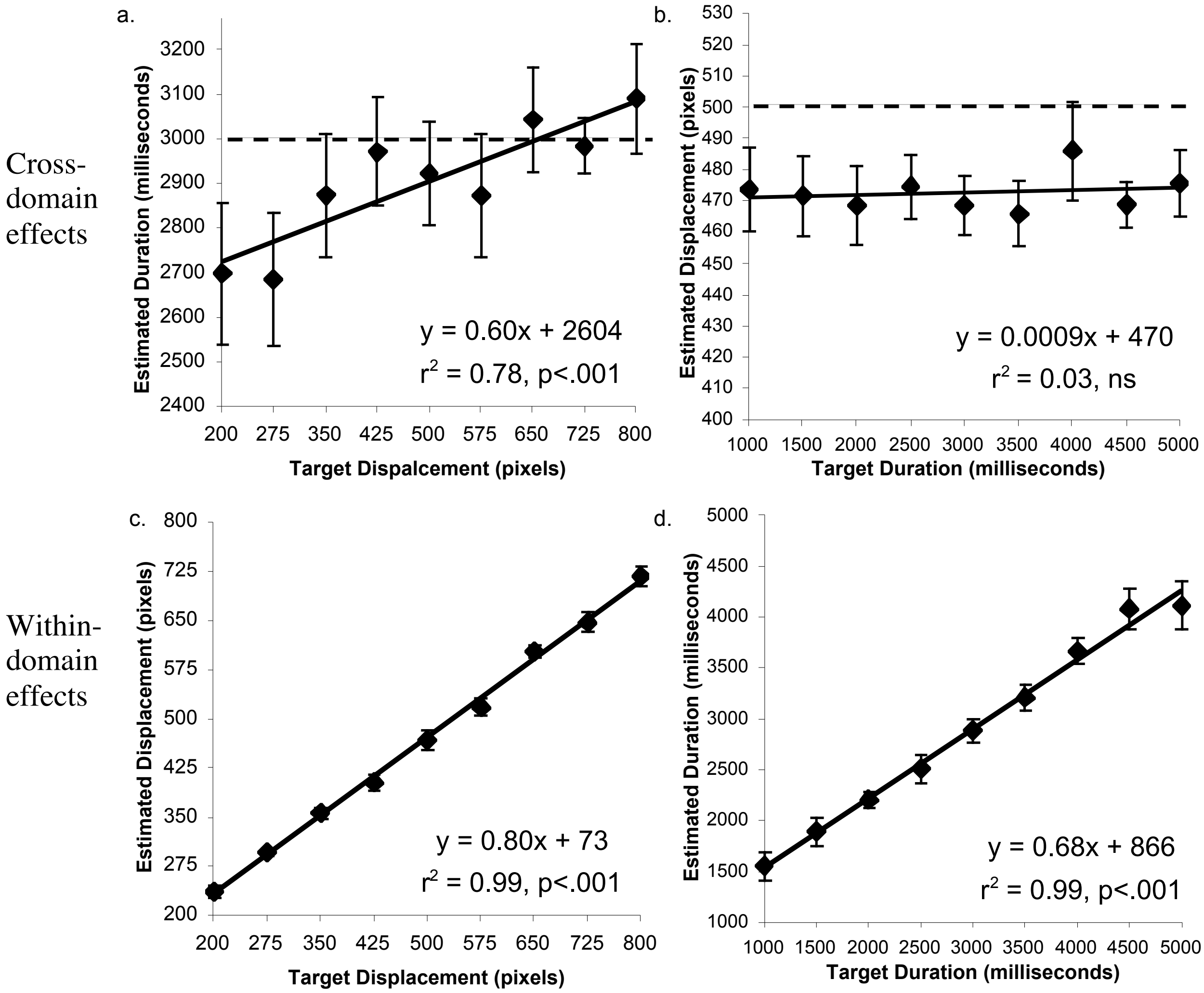

Figure 3. Grand averaged duration and displacement estimates (n=9) for Experiment 3 (Growing Lines, Temporal Frame of Reference). Top: Cross-domain effects. 3a. (left) Effect of displacement on duration estimation. 3b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 3a. and 3b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 3c. (left) Effect of target displacement on estimated displacement. 3d. (right) Effect of target duration on estimated duration.

2.4 Experiment 4: Growing Lines, Concurrent Tone

Would space still influence participants' time estimates if stimulus duration were indexed by something non-spatial? For Experiment 4, a tone of constant frequency and amplitude accompanied each growing line. The tone began sounding when the line started to grow across the screen, and stopped sounding when the line disappeared. Importantly, each tone had a clearly perceptible duration, but no perceptible spatial extent. Thus, stimulus duration was made available to the participant in both the visual and auditory modalities, but stimulus displacement was only available visually.

Materials and Procedure

Stimulus materials and procedures were identical to those used in Experiment 2, with the following addition. A constant tone (260 Hz) accompanied each growing line.

Results and Discussion

Results (figure 4a-d) did not differ significantly from those of the previous experiments. As before, line displacement strongly modulated duration estimations, but line duration did not significantly modulate displacement estimations (figure 7d). Although the tones provided a non-spatial index of line duration, they did not diminish the influence of distance on time estimation.

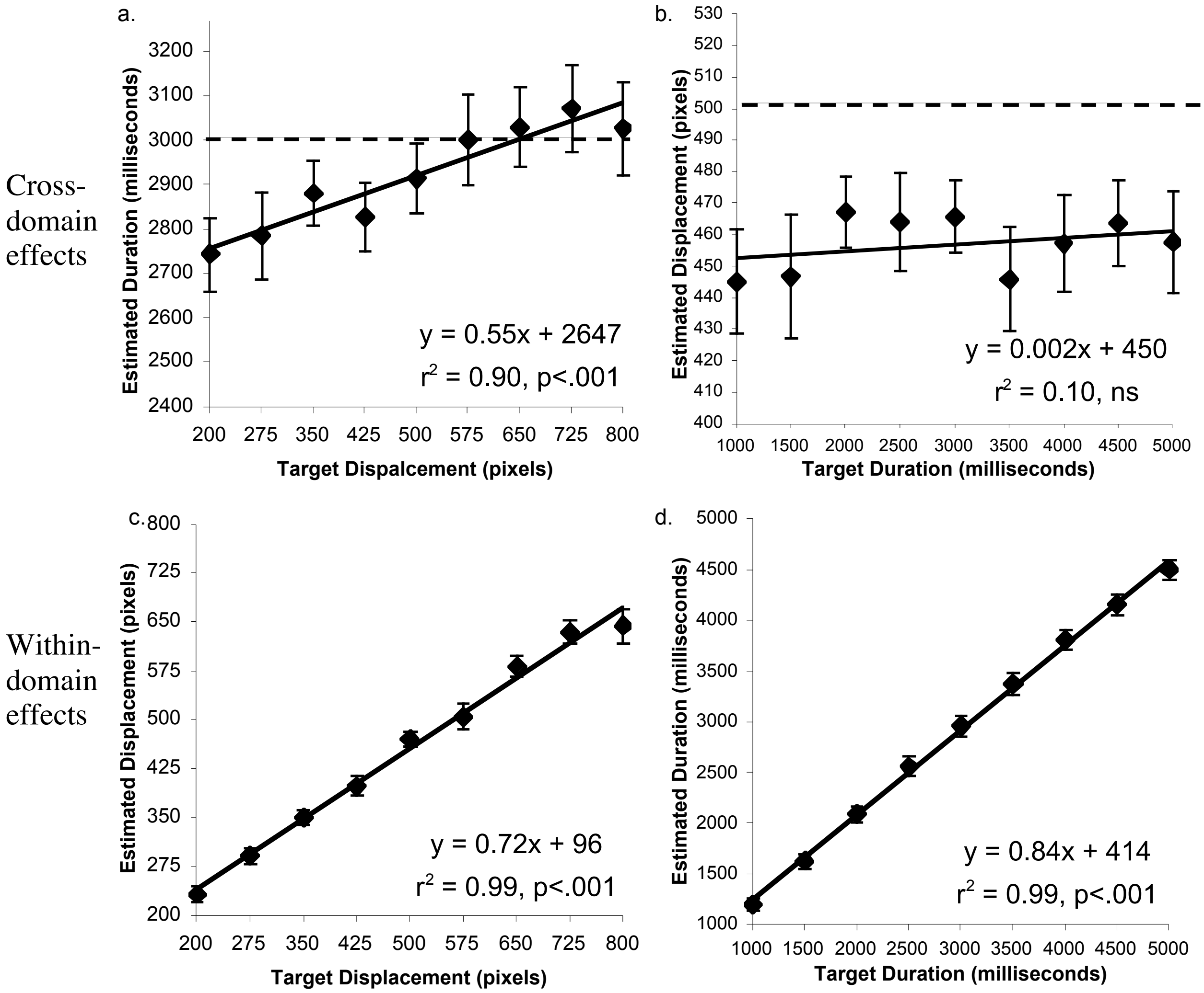

Figure 4. Grand averaged duration and displacement estimates (n=16) for Experiment 4 (Growing Lines, Concurrent Tone). Top: Cross-domain effects. 4a. (left) Effect of displacement on duration estimation. 4b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 4a. and 4b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 4c. (left) Effect of target displacement on estimated displacement. 4d. (right) Effect of target duration on estimated duration.

2.5 Experiment 5: Moving Dot

Was it necessary for participants to see a long line, or would a more abstract experience of spatial displacement also suffice to lengthen their time estimates? For Experiment 5, rather than viewing a growing line, participants saw a moving dot. In the previous experiments, just before each line disappeared participants could see its full spatial extent, from end to end, seemingly at a glance. By contrast, the extent of a moving dot's path could never be seen all at once, rather it had to be imagined: in order to compute the distance that a dot traveled, participants had to retrieve the dot's starting point from memory once its ending point was reached. The spatial and temporal dimensions of the dot stimulus had to be processed similarly in this regard, since it is always the case when we compute the extent of a temporal interval that we must retrieve its starting point from memory once the end of the interval is reached.

Materials and Procedure

Stimulus materials and procedures were identical to those used in Experiment 2, with one exception. Rather than viewing a growing line, subjects viewed a dot (10x10 pixels) that moved horizontally across the midline of the screen, from left to right.

Results and Discussion

Results (figure 5a-d) did not differ significantly from those of the previous experiments. As before, there was a strong and asymmetric cross-dimensional effect of space on time (figure 7e), suggesting that participants' more abstract representations of displacement (i.e., representations of spatial intervals that were never perceived at a glance, but only reconstructed from memory) were sufficient to modulate their duration estimates.

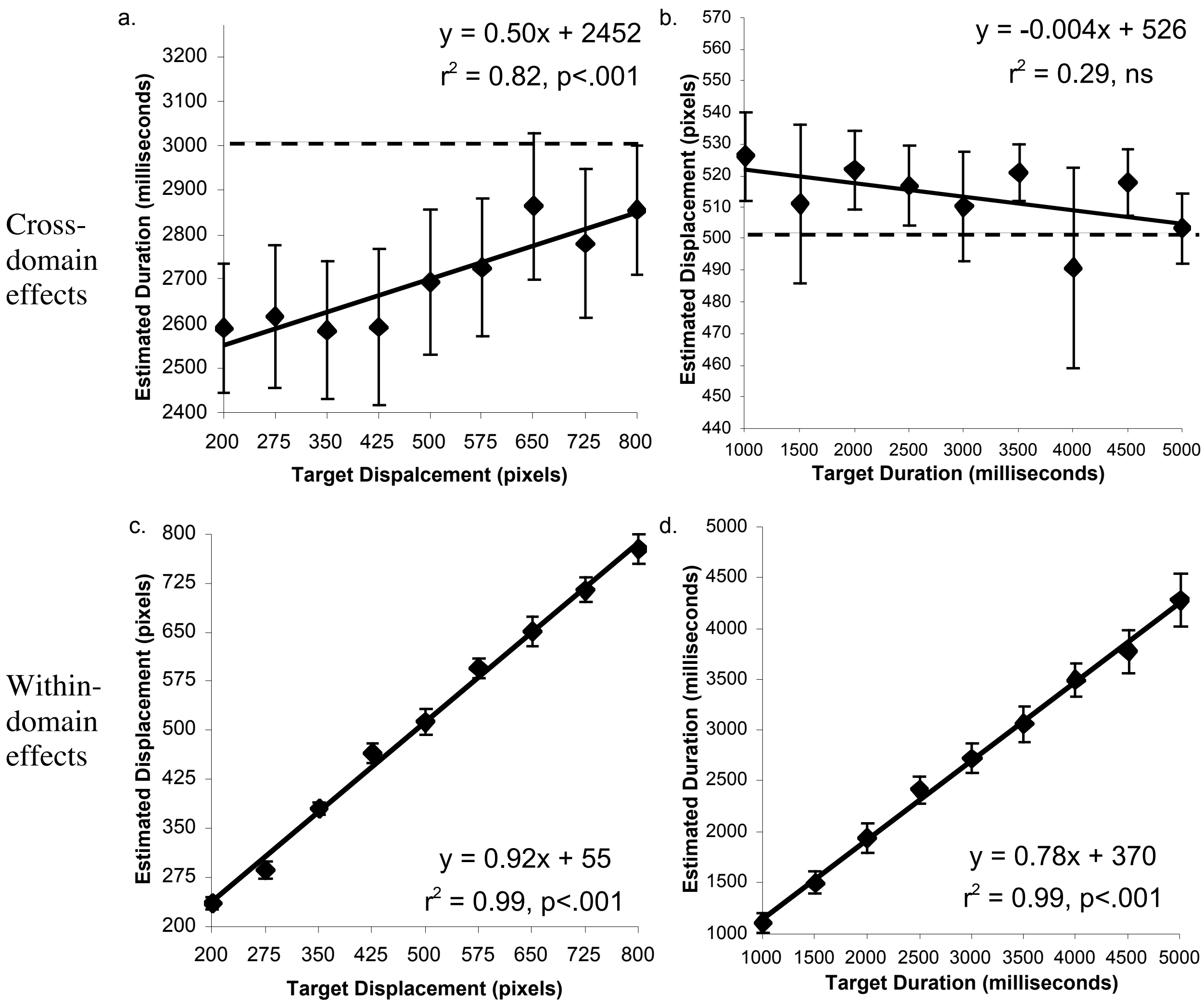

Figure 5. Grand averaged duration and displacement estimates (n=10) for Experiment 5 (Moving Dot). Top: Cross-domain effects. 5a. (left) Effect of displacement on duration estimation. 5b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 5a. and 5b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 5c. (left) Effect of target displacement on estimated displacement. 5d. (right) Effect of target duration on estimated duration.

2.6 Experiment 6: Stationary Lines

How do people come to depend, in part, on distance to make time estimations? One possibility is that they are sensitive to the correlation of displacement and duration in their everyday experience with moving objects: any change in an object's position is necessarily accompanied by a change in time. People may rely on spatial change heuristically as an index of temporal change, perhaps because spatial representations are more durable than temporal representations, or because motion through space is more directly perceptible than 'motion' through time. Experiments 1-5 used moving stimuli. Would the asymmetric relationship between space and time still be found if static stimuli were used? In Experiment 6, participants viewed stationary lines and estimated either their displacement from end to end or the amount of time they remained on the screen, as in previous experiments.

Materials and Procedure

Stimulus materials and procedures were identical to those used in Experiment 2, with the following exception. Rather than viewing growing lines, participants viewed stationary lines of various displacements, which remained on the screen for various durations, according to the parameters used in Experiment 2.

Results and Discussion

Results (figures 6a-d) showed essentially the same pattern of results found in all previous experiments. Duration estimates were strongly and asymmetrically dependent on stimulus displacement (figure 7f). This finding rules out the possibility that motion or speed was principally responsible for the results of the previous experiments. Nevertheless, motion or speed may have contributed to the cross-dimensional interference observed in Experiments

1-6. The strength of the effect of displacement on duration estimation was contrasted for the stationary line experiment (Experiment 6) and the comparable growing line experiment (Experiment 2). Although no significant difference was found between the correlations ($r_{\text{growing lines}} - r_{\text{stationary lines}} = 0.11$; $z = 1.19$, *ns*), a difference in slopes indicated that the effect of distance on time estimation was stronger in Experiment 2 than in Experiment 6. Results of a 2-way mixed ANOVA with Task as a between subjects factor (growing lines, stationary lines) and Cross-Dimensional Interference Type as a within subject factor (standardized slope of the effect of space on time, standardized slope of the effect of time on space) showed a significant main effect of Cross-Dimensional Interference Type ($\underline{M}_{\text{effect of time on space}} = 0.007$, SD = 0.06; $\underline{M}_{\text{effect of space on time}} = 0.07$, SD = 0.07; $F(1, 26) = 18.3$, $p < .001$), a marginally significant main effect of task ($\underline{M}_{\text{growing lines}} = 0.06$, SD = 0.09; $\underline{M}_{\text{stationary lines}} = 0.03$, SD = 0.06; $F(1, 26) = 3.5$, $p < .07$), and importantly, a significant interaction of Task and Cross-Dimensional Interference Type ($F(1, 26) = 5.3$, $p < 0.03$). Post-hoc t-tests revealed that the standardized slope of the effect of space on time estimation was significantly greater for the growing line task ($\underline{M} = 0.12$, SD = 0.07) than for the stationary line task ($\underline{M} = 0.05$, SD = 0.06; difference of slopes = 0.07, df = 26, $t = 2.8$, $p<0.01$, two-tailed). By contrast, the standardized slope of the effect of time on space did not differ between the growing line task ($\underline{M} = -0.0004$, SD = 0.06) and the stationary line task ($\underline{M} = 0.01$, SD = 0.06; difference of slopes = 0.009, df = 26, $t = 0.42$, ns).

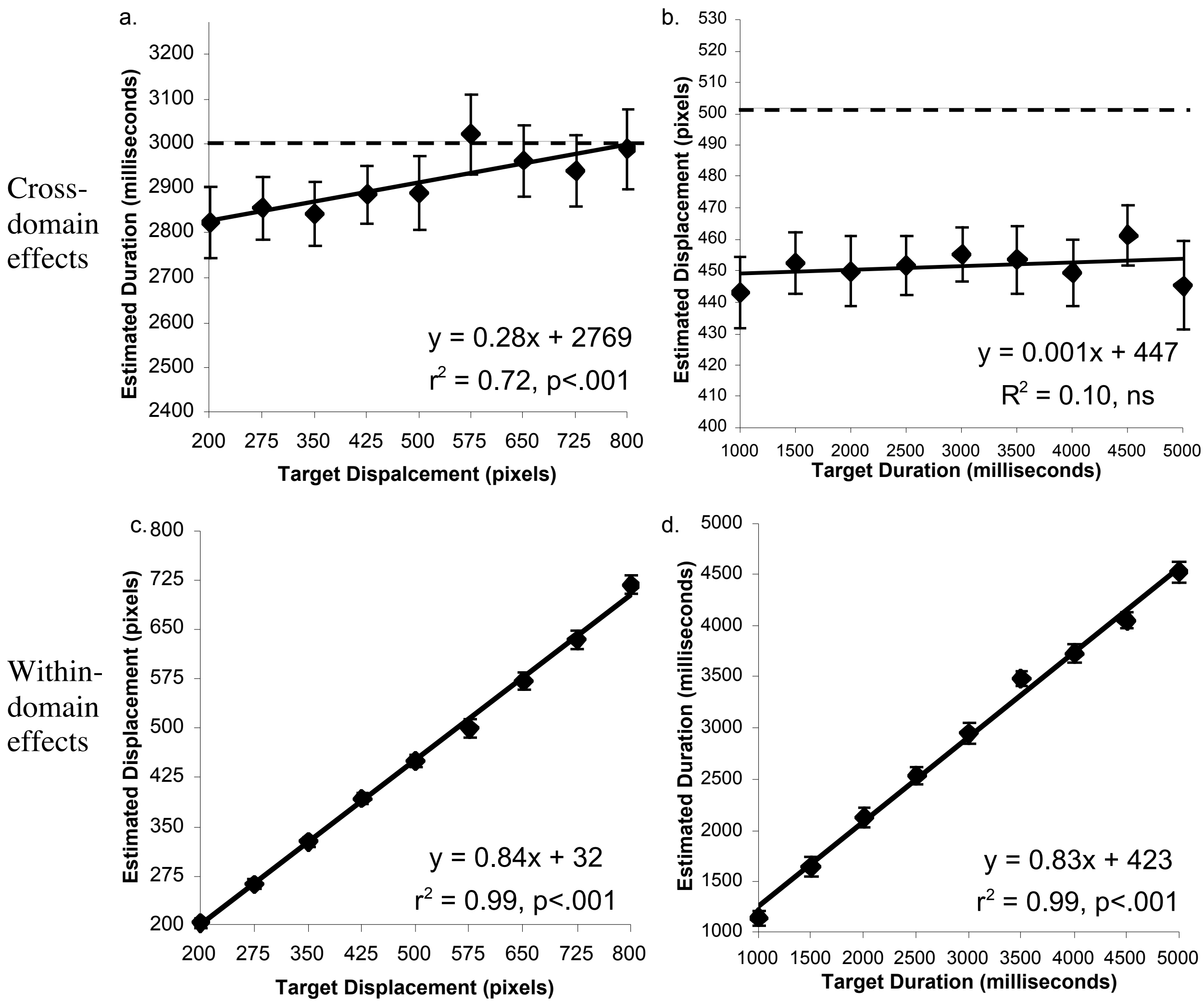

Figure 6. Grand averaged duration and displacement estimates (n=19) for Experiment 6 (Stationary Lines). Top: Cross-domain effects. 6a. (left) Effect of displacement on duration estimation. 6b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 6a. and 6b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 6c. (left) Effect of target displacement on estimated displacement. 6d. (right) Effect of target duration on estimated duration.

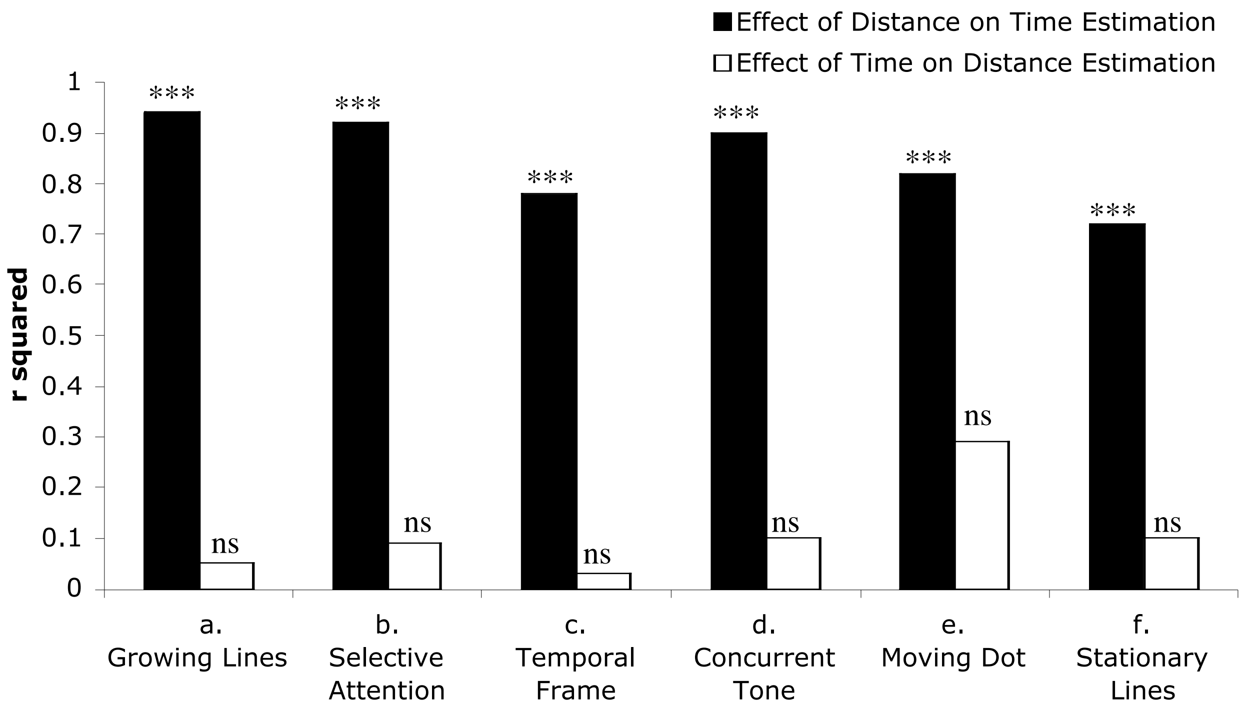

Figure 7. Summary of cross-dimensional interference effects. (*** indicates correlation was significant at $p<.001$). The effect of distance on time estimation was significantly greater than the effect of time on distance estimation for all experiments. (7a, Growing lines: difference of correlations = 0.75; $z = 3.24$, $p <.001$. 7b, Growing lines, selective attention: difference of correlations = 0.66; $z = 2.84$, $p < .003$. 7c, Growing lines, temporal frame of reference: difference of correlations = 0.71; $z =2.09$, $p <.02$. 7d, Growing lines, concurrent tone: difference of correlations =0.63; $z = 2.60$, $p <.005$. 7e, Moving dot: difference of correlations = 1.45; $z = 3.69$, $p <.001$. 7f, Stationary lines: difference of correlations = 0.54; $z = 1.62$, $p <.05$.)

2.7 Experiment 7: Orientation and direction of temporal thinking

In experiments 1-6, all stimuli were horizontal, and all moving stimuli progressed from left to right: the direction of reading and writing in English, and the direction of increase in many graphs. Could the results of Experiments 1-6 be an artifact of literacy? Would the same relation between distance and time be found if stimuli traveled from right to left, or if lines grew upward or downward? For Experiments 7a-d, English speaking participants performed one of four distance/time estimation tasks in which lines grew either rightward, leftward, upward, or downward.

Previous research suggests that spatial schemas underlying some kinds of time representations are specified in both direction and orientation. Boroditsky's experiments reviewed in section 1.3 show that English speakers often talk -- and think -- about time as if it flows in a particular direction, and along a horizontal rather than a vertical axis (e.g., moving the meeting *forward*; pushing the deadline *back*) (Boroditsky, 2000, 2001; Clark, 1973). The direction in which time flows appears to be determined, in part, by the speaker's orthography. Tversky, Kugelmass, and Winter (1991) found that participants spontaneously mapped a series of events (i.e., breakfast, lunch, dinner) onto a horizontal line that was directed rightward if their first language was English, but directed leftward if their first language was Arabic, which is written right-to-left.

Although these studies are informative about representations of temporal succession, little is known about the specificity of spatial schemas supporting representations of duration. Unlike linguistic metaphors for succession, linear spatial metaphors for duration do not seem to imply a specific orientation, and may not encode direction, at all. When English speakers borrow distance terms like *long* and *short* to talk about events, it is not clear whether they are importing horizontal or vertical spatial schemas into the domain of time. In spatial contexts, *long* and *short* often describe horizontal extent (e.g., *a long boulevard*, *a short driveway*), but these words also commonly refer to vertical extent (e.g., *a short skirt*, *a long braid*, *a short drop*, *a long way up*, etc.) The orientation of *long* and *short* is flexible, but can be specified in spatial contexts. Beyond being flexible, the direction of *length* or *shortness* seems wholly unspecified, in many cases. Upon hearing a sentence like "the rope was long," a listener might not know the orientation of the rope, which could be running along the edge of a driveway or dangling from a cliff. Upon hearing "the rope on the flagpole was long," the listener may assume that the rope is oriented vertically, but its length is still not specified in terms of direction: the rope has the same length whether measured from bottom to top or from top to bottom[5]. Thus, a distance metaphor "like the meeting was long" appears to import some aspects of spatial experience into the domain of time, but not others: *extent* and *dimensionality* are specified (i.e., spatial distance is unidimensional), whereas *orientation* and *direction* are unspecified. The flexibility of distance metaphors for duration seems consistent with the relation between distance and time in our physical experience with moving objects: as an object travels farther more time passes, regardless of the direction or orientation in which the object is traveling.

Contrasting predictions regarding the effect of distance on time estimation in Experiments 7a-d can be generated based on English speakers' experience with written language as opposed to their experience with distance metaphors in spoken language and moving objects in the environment. It is possible that people learn an association between progress in distance and time through the habit of scanning across the printed page, and that this association contributed to the cross-dimensional interference reported in Experiments 1-6. If the habit of reading and writing from left to right was responsible for

[5] In other cases such as *the skirt was short*, direction may be more strongly implied (i.e,. the skirt goes a short distance down from the waist, not up from the hem).

the positive correlation between distance and time estimation in the previous experiments, then this correlation should be weaker (or perhaps even negative) when lines grow in the opposite direction. This predicts a stronger effect of distance on time estimation in Experiment 7a (Rightward Growing Lines) than in 7b (Leftward Growing Lines). When we read and write in English, we progress not only from the left to the right but also from the top to the bottom of the page (likewise when we use calendars, PDAs, train timetables, etc.) If these habits produce associations between downward displacement and time, then a stronger effect of distance on time estimation should be found in Experiment 7d (Downward Growing Lines) than in 7c (Upward Growing Lines). By contrast, if English speakers' mental representations of duration are shaped by distance metaphors in language or by physical experience with moving objects, then no differences are predicted in the effect of distance on time estimation across Experiments 7a-d.

Materials and Procedure

For Experiment 7a (Rightward Growing Lines), stimulus materials and procedures were identical to those used in Experiment 2, with the following exceptions. The range of the nine line displacements was reduced from 200-800 pixels to 100-500 pixels, increasing in 50 pixel increments. This change was necessary in order for lines to fit on the 1024 X 768 pixel screen when rotated in vertical versions of the experiment. Experiments 7b (Leftward Growing Lines), 7c (Upward Growing Lines), and 7d (Downward Growing Lines) were identical to the 7a, except for the direction and orientation of the growing lines. For all four experiments, a square box (700 x 700 pixels) framed the portion of the screen where stimuli appeared, to minimize any influence of the asymmetric rectangular shape of the computer monitor on perception of horizontal versus vertical lines.

Although the design of Experiments 7a-d included nine durations crossed with nine displacements as in previous experiments, for the current experiments only the middle five durations and displacements were analyzed as target trials. The two highest and lowest durations and displacements were treated as filler trials. While it is common in magnitude estimation tasks for the endpoints to be trimmed in order to avoid experimental artifacts, this decision was also motivated by: (a) the observation that in previous experiments the strongest linear relations between distance and time were found in the middle of the range of stimuli, (b) concern that the longest and shortest lines (in space and time) were most amenable to verbal labeling, whereas the middle stimuli were least likely to be covertly labeled as *long* or *short*, (c) to further reduce the possibility that the asymmetric shape of the monitor could influence judgments of horizontal versus vertical stimuli.

Results and Discussion

Results of Experiments 7a-d are presented in figures 8a-d, 9a-d, 10a-d, and 11a-d. A similar pattern of results was found across all four experiments: there was a significant positive correlation between target displacement and estimated duration, but no significant correlation between target duration and estimated displacement. Although the effect of distance on time estimation appeared somewhat stronger in Experiment 7a (Rightward Growing Lines) than in Experiments 7b-d, pairwise comparisons revealed no differences among the correlations. Likewise, a one-way ANOVA comparing the slopes of the effect of distance on time estimation showed no significant differences as a function of the direction or orientation of the growing lines (figure 12). Post-hoc tests were not mandated due to the nonsignificant ANOVA results, but pairwise comparisons of slopes were conducted nevertheless, to confirm that there were no significant differences across experiments 7a-d.

Two further analyses tested for effects of the direction and orientation of growing lines on time estimation. First, to test for an effect of orientation independent of direction, data were pooled for the horizontal line experiments, 7a (Rightward) and 7b (Leftward), and compared with the pooled data from the vertical line experiments, 7c (Upward) and 7d (Downward). No significant difference was found between the slopes of the effect of distance on time estimation for the horizontal experiments ($\underline{M}_{horizontal}$ = 1.69, SD = 1.72) versus vertical experiments ($\underline{M}_{vertical}$ = 1.24, SD = 1.28; difference of slopes = 0.45, t(51) = 1.09, ns). Second, to test for the influence of English orthography on time estimation, data were pooled for the orthography-consistent experiments, 7a (Rightward) and 7d (Downward), and compared with the pooled data from the orthography-inconsistent experiments, 7b (Leftward) and 7c (Upward). No significant difference was found ($\underline{M}_{Orthography\text{-}consistent}$= 1.45, SD = 1.44; $\underline{M}_{Orthography\text{-}inconsistent}$= 1.48, SD = 1.64; difference of slopes = 0.03, t(51) = -0.05, ns). These results mediate against that possibility that the association between distance and time estimation observed in Experiments 1-7 was created by the habit of reading and writing. Rather, results are consistent with associations between distance and time found in linear spatial metaphors for duration (as opposed to succession) in English, and with the relation between distance and duration in our experience of moving objects.

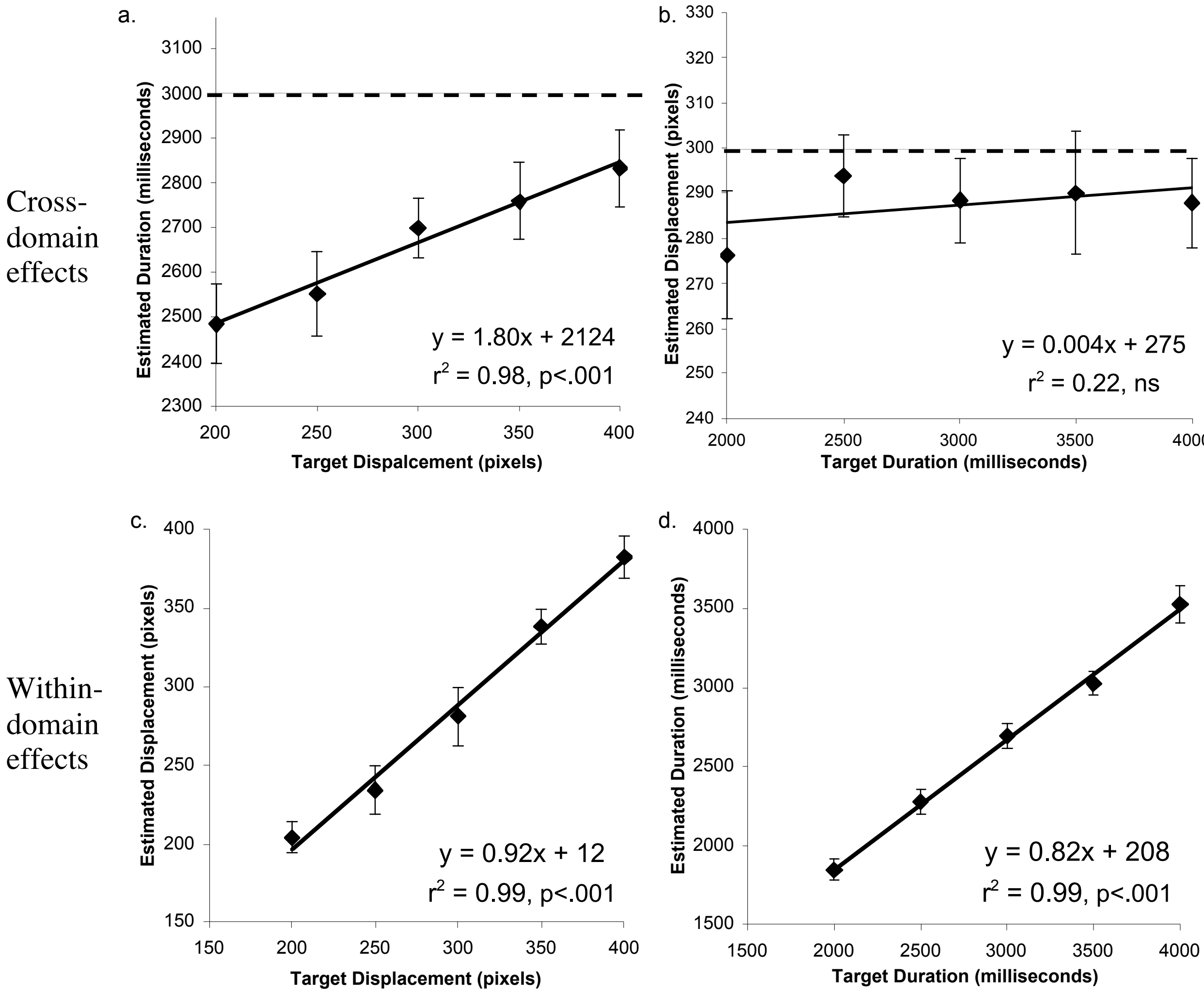

Figure 8. Grand averaged duration and displacement estimates (n=17) for Experiment 7a (Rightward Growing Lines). Top: Cross-domain effects. 8a. (left) Effect of displacement on duration estimation. 8b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 8a. and 8b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 8c. (left) Effect of target displacement on estimated displacement. 8d. (right) Effect of target duration on estimated duration.

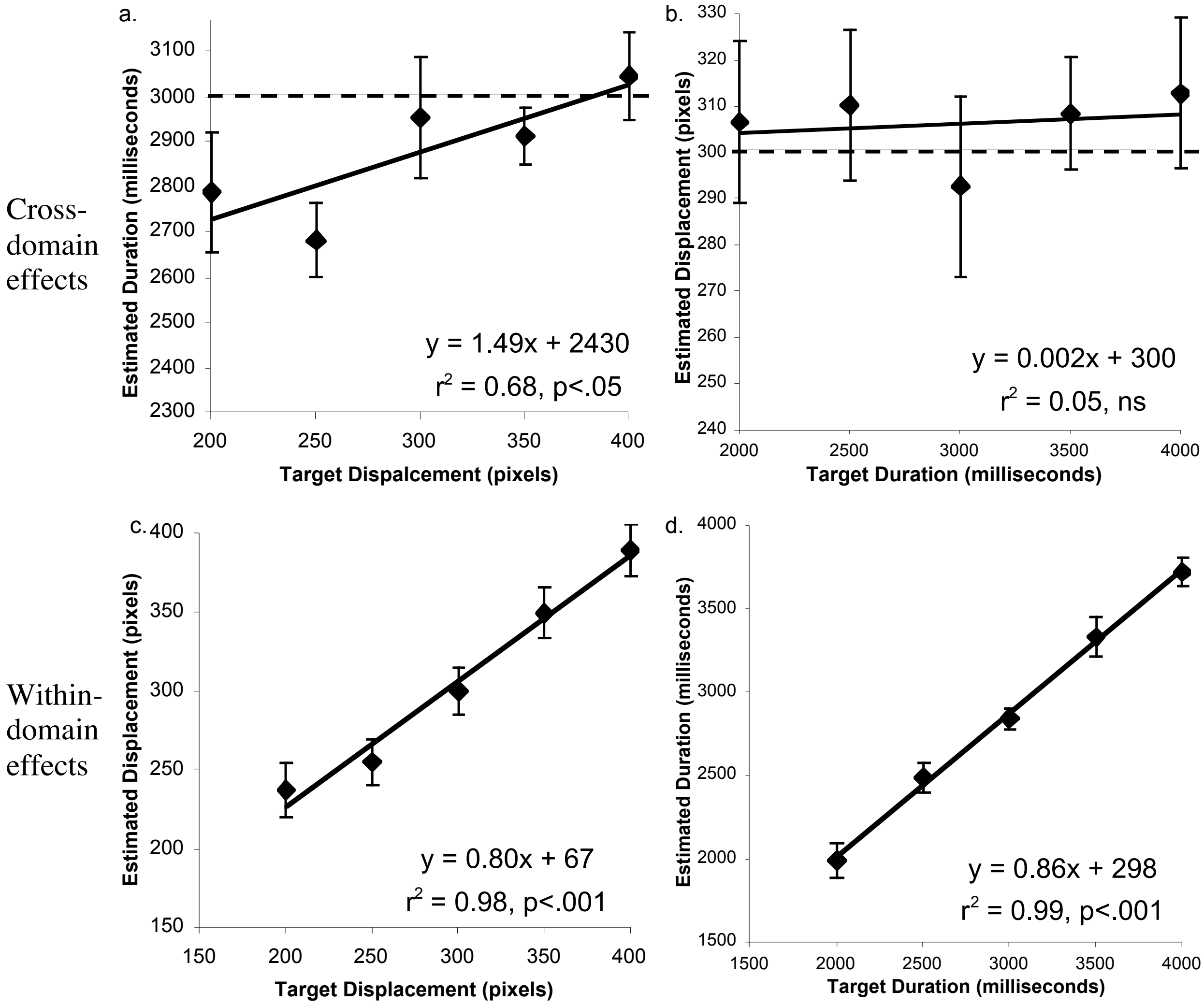

Figure 9. Grand averaged duration and displacement estimates (n=9) for Experiment 7b (Leftward Growing Lines). Top: Cross-domain effects. 9a. (left) Effect of displacement on duration estimation. 9b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 9a. and 9b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 9c. (left) Effect of target displacement on estimated displacement. 9d. (right) Effect of target duration on estimated duration.

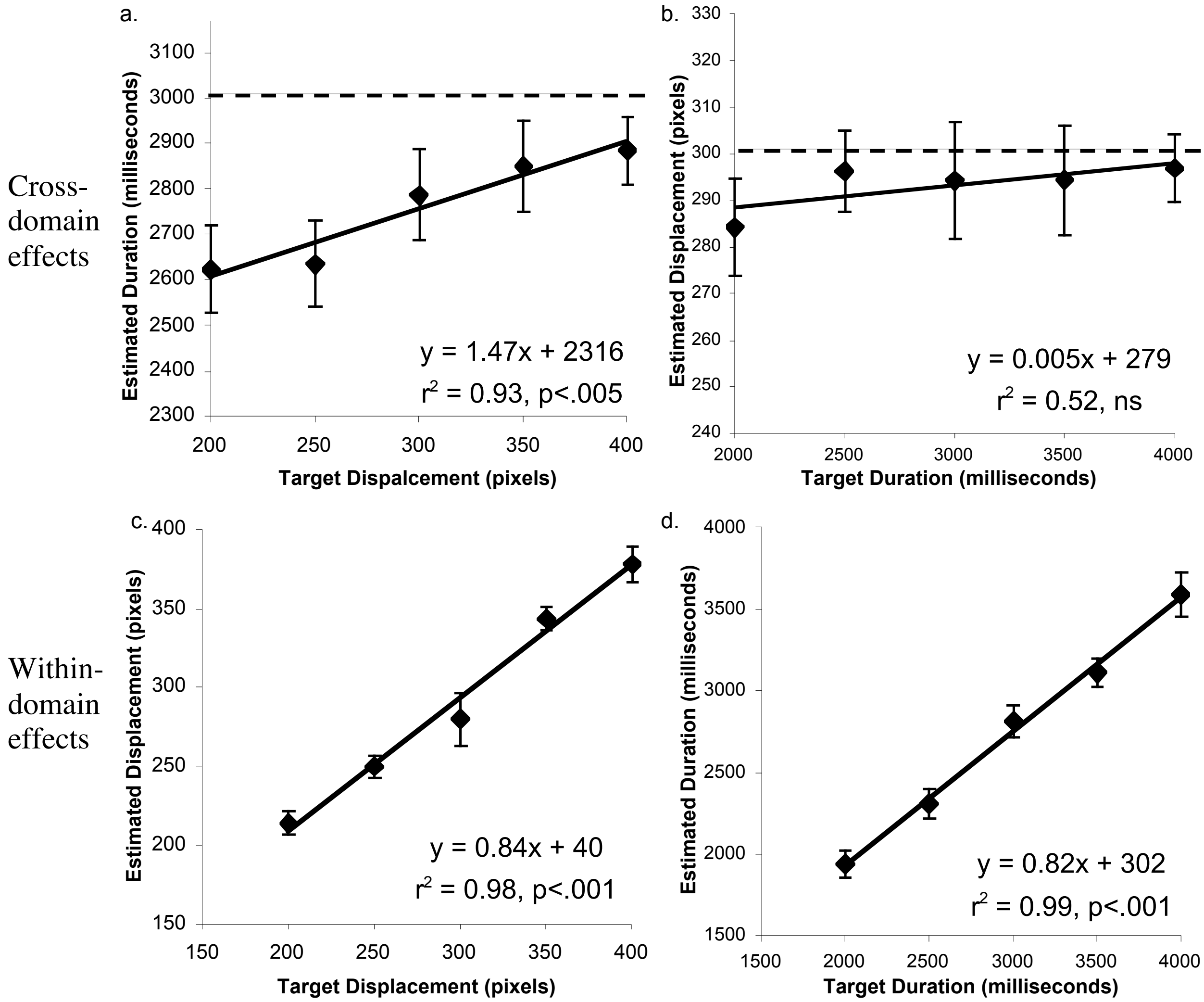

Figure 10. Grand averaged duration and displacement estimates (n=14) for Experiment 7c (Upward Growing Lines). Top: Cross-domain effects. 10a. (left) Effect of displacement on duration estimation. 10b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 10a. and 10b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 10c. (left) Effect of target displacement on estimated displacement. 10d. (right) Effect of target duration on estimated duration.

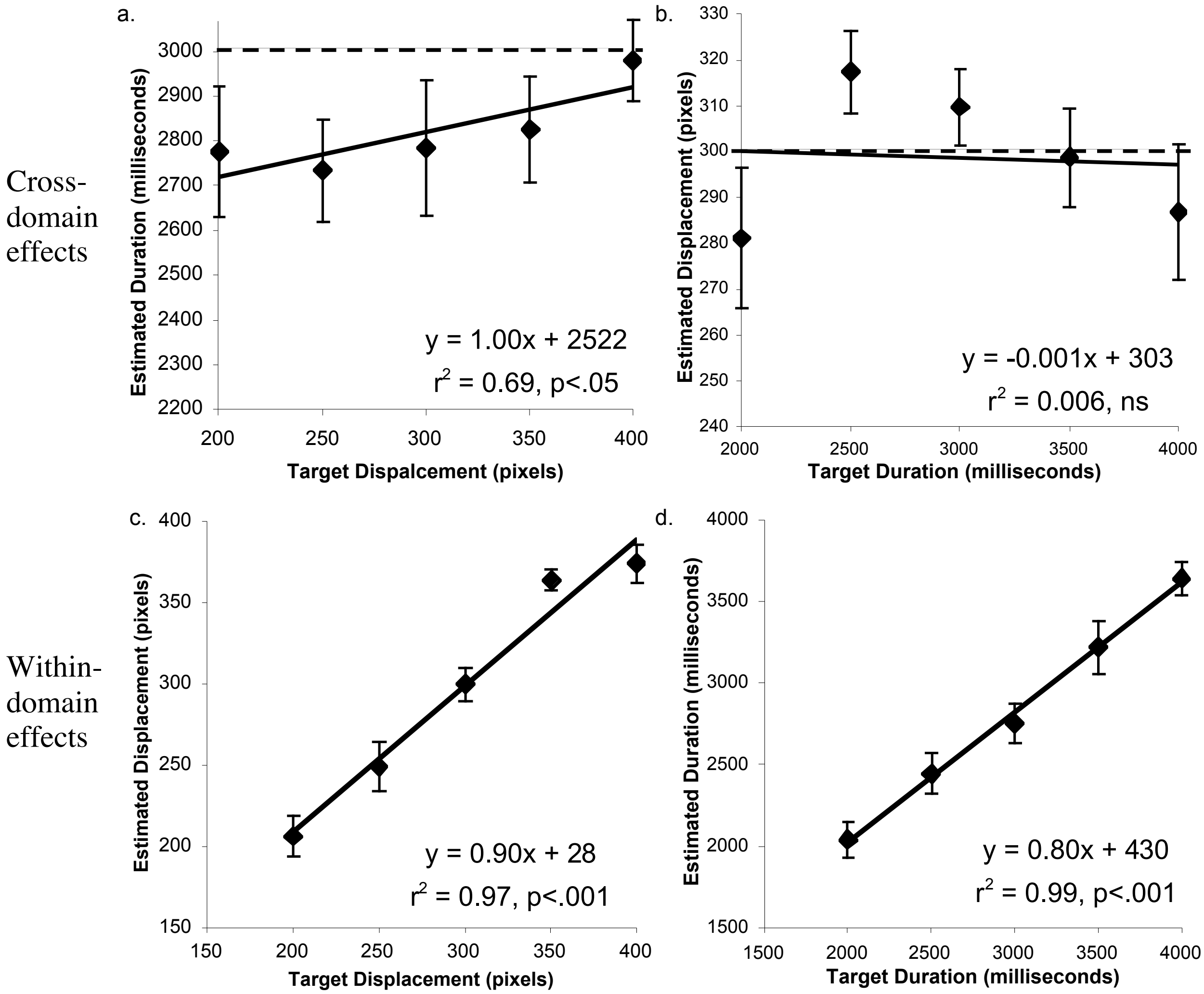

Figure 11. Grand averaged duration and displacement estimates (n=13) for Experiment 7d (Downward Growing Lines). Top: Cross-domain effects. 11a. (left) Effect of displacement on duration estimation. 11b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 11a. and 11b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 11c. (left) Effect of target displacement on estimated displacement. 11d. (right) Effect of target duration on estimated duration.

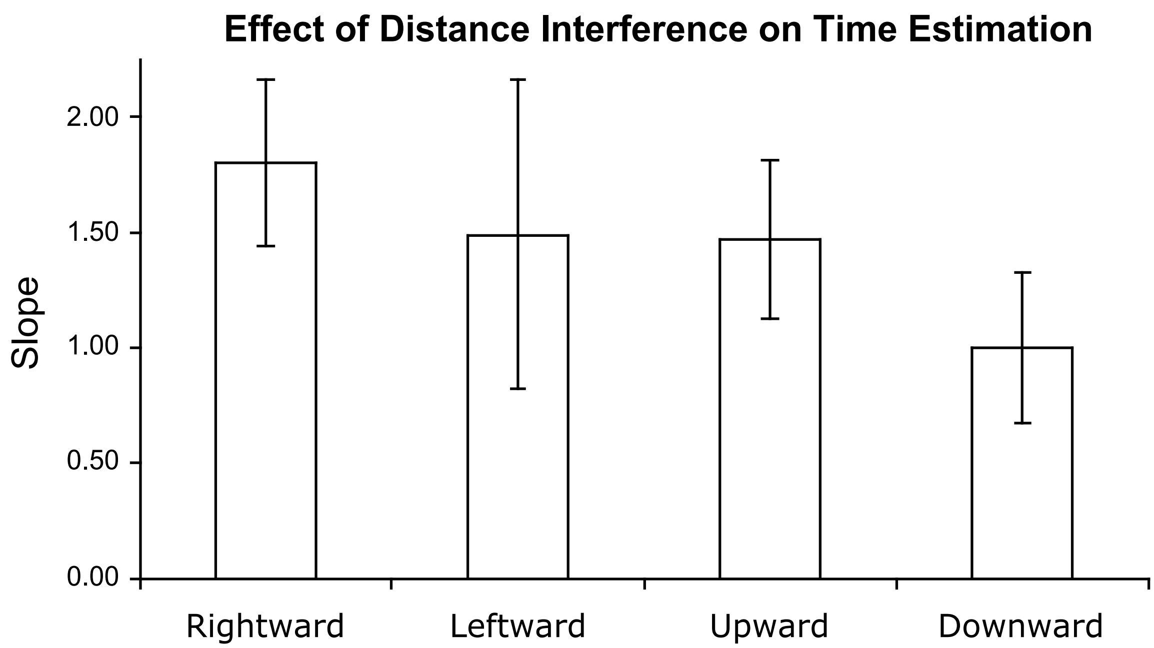

Figure 12. Summary of the effects of distance interference on duration estimation for experiments 7a-d. The slope of the effect of target displacement on estimated duration did not differ significantly as a function of the orientation or direction of the growing lines ($F(3,49) = .68$, $p = .57$).

Chapter 3: Does language shape the way we think about time?

The first set of experiments support the *Deep View* of language-thought relations by showing that temporal representations depend, in part, on spatial representations, as predicted by metaphors in English -- even when people are performing low-level, nonlinguistic psychophysical tasks (see Table 1, number i). However, it is not clear from these data whether linguistic metaphors merely reflect English speakers' underlying nonlinguistic representations of time, or whether language also shapes those representations. According to the *Shallow View,* it is possible that speakers of a language with different duration metaphors would nevertheless perform similarly to English speakers on nonlinguistic tasks. Thus, these experiments leave the Whorfian question unaddressed. In Whorf's words:

> Are our own concepts of 'time,' 'space,' and 'matter' given in substantially the same form by experience to all men, or are they in part conditioned by the structure of particular languages?" (1939/2000, pg. 138.)

This question, posed over half a century ago, remains the subject of renewed interest and debate. Does language shape thought? The answer *yes* would call for a reexamination of the 'universalist' assumption that has guided Cognitive Science for decades, according to which nonlinguistic concepts are formed independently of the words that name them, and are invariant across languages and cultures (Fodor, 1975; Pinker, 1994, Papafragou, Massey, & Gleitman, 2002)[6]. Although members of the general public may be quick to believe that people who talk differently also think differently (ask anyone about the Eskimos and their words for snow), many linguists and psychologists remain unconvinced.[7]

---

[6] The 'universalist' position is often attributed to Chomsky (1975), but has been articulated more recently by Pinker (1994) and by Lila Gleitman and colleagues (Papafragou, Massey, & Gleitman, 2002; Snedeker & Gleitman, 2004). The *Shallow View* proposed here can be considered a variety of the universalist view that maintains its validity despite recent psycholinguisitc evidence supporting the Whorfian hypothesis (e.g., Boroditsky, 2001).

[7] See Pinker, 1994, chapter 3, for a review of evidence against the Whorfian Hypothesis.

Skepticism about some Whorfian claims has been well founded. Two crucial kinds of evidence have been missing from previous inquiries into relations between language and thought: objectively evaluable linguistic data, and language-independent psychological data. A notorious fallacy, attributable in part to Whorf, illustrates the need for methodological rigor. Whorf (1939/2000) argued that Eskimos must conceive of snow differently than English speakers because the Eskimo lexicon contains multiple words that distinguish different types of snow, whereas English has only one word to describe all types. The exact number of snow words the Eskimos were purported to have is not clear. (This number has now been inflated by the popular press to as many as four-hundred.) According to a Western Greenlandic Eskimo dictionary published in Whorf's time, however, Eskimos may have had as few as two distinct words for snow (Pullum, 1991).

Setting aside Whorf's imprecision and the media's exaggeration, there remain two problems with Whorf's argument, which are evident in much subsequent 'Language and Thought' research, as well. First, although Whorf asserted an objective difference between Eskimo and English snow vocabularies, his comparative linguistic data were subjective and unfalsifiable: it is a matter of opinion whether any cross-linguistic difference in the number of snow words existed. As Geoffrey Pullum (1991) points out, English could also be argued to have multiple terms for snow in its various manifestations: *slush, sleet, powder, granular*, *blizzard, avalanche, etc*. The problem of unfalsifiability could be addressed if cross-linguistic differences could be demonstrated empirically, and ideally, if the magnitude of the differences could be quantified.

A second problem with Whorf's argument (and others like it in the contemporary Cognitive Linguistics literature) is that it uses purely linguistic data to motivate inferences about nonlinguistic thinking. Steven Pinker illustrates the resulting circularity of Whorf's claim in this parody of his logic:

> [They] speak differently so they must think differently.
> How do we know that they think differently?
> Just listen to the way they speak! (Pinker, 1994, pg. 61).

Such circularity would be escaped if nonlinguistic evidence could be produced to show that two groups of speakers who talk differently also think differently in corresponding ways.

Does language shape the way we think about time? Experiment 8 uncovered previously unexplored cross-linguistic differences in spatial metaphors for duration. Experiments 9 and 10 tested whether these linguistic differences correspond to differences in speakers' low-level, nonlinguistic time representations. Experiment 11 evaluated a causal role for language in shaping time representations[8].

3.1 1-dimensionsal and 3-dimensional spatial metaphors for time

Literature on how time can be expressed (and by hypothesis conceptualized) in terms space has focused principally on linear spatial metaphors. But is time necessarily conceptualized in terms of unidimensional space? Some theorists have suggested so (Clark, 1973, Gentner, 2001), and while this may be true regarding temporal succession, linguistic metaphors suggest an alternative spatialization for duration. English speakers not only describe time as a line, they also talk about *oceans of time, saving time in a bottle*, and liken the 'days of their lives' to *sand through the hourglass*: apparently mapping time onto quantities accumulating in three dimensions (i.e., volume).

Experiment 8 compared the use of 'time as distance' and 'time as quantity' metaphors across four languages. Every language examined uses both distance and quantity metaphors, but their relative prevalence and productivity appear to vary markedly. In English, it is natural to talk about *a long time*, borrowing the structure and vocabulary of a spatial expression like *a long rope*. Yet in Spanish, the direct translation of 'long time',

---

[8] A preliminary report on these experiments appeared as Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, S., Santiago-Diaz, I., Fotokopoulu, O., Pita, R., & Gil, D., (2004). *How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish*. Proceedings of the 26th Annual Conference of the Cognitive Sciences Society, Chicago, IL.

largo tiempo, sounds awkward to speakers of most dialects[9]. *Mucho tiempo*, which means 'much time', is preferred .

In Greek, the words *makris* and *kontos* are the literal equivalents of the English spatial terms *long* and *short*. They can be used in spatial contexts much the way *long* and *short* are used in English (e.g., *ena* <u>*makry*</u> *skoini* means 'a <u>long</u> rope'). In temporal contexts, however, *makris* and *kontos* are dispreferred in instances where *long* and *short* would be used naturally in English. It would be unnatural to translate *a* <u>*long*</u> *meeting* literally as *mia* <u>*makria*</u> *synantisi*. Rather than using distance terms, Greek speakers typically indicate that an event lasted a long time using *megalos*, which in spatial contexts means physically 'large' (e.g., a big building), or using *poli*, which in spatial contexts means 'much' in physical quantity (e.g., much water). Compare how English and Greek typically modify the duration of the following events (literal translations in parentheses):

1e. long night
1g. megali nychta (*big night*)

2e. long relationship
2g. megali schesi (*big relationship*)

3e. long party
3g. parti pou kratise poli (*party that lasts much*)

4e. long meeting
4g. synantisi pou diekese poli (*meeting that lasts much*)

In examples 1g. and 2g., the literal translations might surprise an English speaker, for whom *big night* is likely to mean 'an exciting night', and *big relationship* 'an important relationship'. For Greek speakers, however, these phrases communicate duration, expressing time not in terms of unidimensional space, but rather in terms of physical quantity (i.e., three-dimensional space).

---

[9] Native speakers of European and South American Spanish report that *largo tiempo* is only used in poetic contexts (e.g., the Peruvian national anthem) to mean *throughout the length of history*. By contrast, some bilingual North American Spanish speakers report that *largo tiempo* can be used colloquially, much like *long time*, perhaps because the construction is imported from English.

Materials and Procedure

To quantify the relative prevalence of distance and quantity metaphors for duration across languages, the most natural phrases expressing the ideas ‘a long time’ and ‘much time’ were elicited from native speakers of English, Indonesian, Greek, and Spanish. The frequencies of these expressions were compared in a very large multilingual text corpus: www.google.com. Each expression was entered as a search term. Google’s language tools were used to find exact matches for each expression, and to restrict the search to web pages written only in the appropriate languages.

Results and Discussion

The number of google ‘hits’ for each expression was tabulated (table 2), and the proportion of distance hits and quantity hits was calculated for each pair of expressions, as a measure of their relative frequency (figure 13). Results showed that in English and Indonesian, distance metaphors were dramatically more frequent than quantity metaphors. The opposite pattern was found in Greek and Spanish. A Chi-Square test showed that the distribution of distance and quantity metaphors varied significantly across languages (X^2=8.5x10^5,df=3, $p<0.001$). These findings corroborate native speakers’ intuitions for each language.

Since this linguistic phenomenon has not been reported previously, results of the Google search were corroborated using a questionnaire study investigating the use of distance and quantity terms to modify event duration in English, Indonesian, Greek, and Spanish (for summary, see Appendix 1).

Table 2.

Frequencies of distance and quantity metaphors for time across languages.

Language	Distance expression	Instances	Quantity expression	Instances
English	*long time*	9,490,000	*much time*	3,440,000
Indonesian	*waktu panjang*	1250	*waktu banyak*	490
Greek	*makry kroniko diatstima*	10,3000	*poli ora*	41,800
Spanish	*largo tiempo*	44,300	*mucho tiempo*	374,000

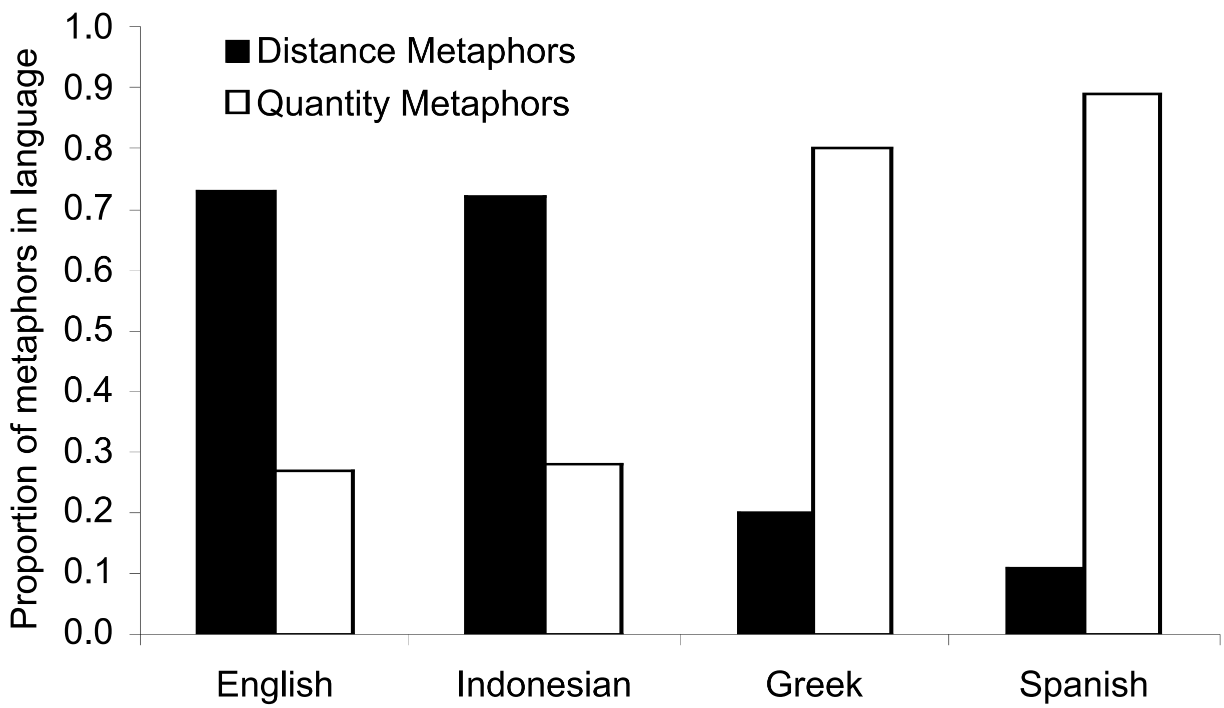

Figure 13. Results of Experiment 8. Black bars indicate the proportion Google 'hits' for expressions meaning *long time*, and white bars for expressions meaning *much time*.

3.2 Do people who talk differently think differently?

Do people who use different spatiotemporal metaphors think about time differently – even when they're not using language? Experiments 9 and 10 explored the possibility that speakers who preferentially use distance metaphors also tend to co-opt linear spatial representations to understand duration, whereas speakers who preferentially use quantity metaphors tend to co-opt 3-dimensional spatial representations. Speakers of the four languages surveyed in Experiment 8 performed a pair of non-linguistic psychophysical tasks, which required them to estimate duration while overcoming different kinds of spatial interference (i.e., distance or volume). If people's conception of time is substantially the same universally irrespective of the language they speak, as suggested by the *Shallow View*, then performance on these tasks should not differ between language groups. On the *Deep View*, however, it was predicted that performance should vary in ways that parallel participants' language-particular metaphors.

The 'distance interference' task was identical to the rightward horizontal growing line task, 7a. English participants in the previous growing line studies may have suffered interference of distance on duration estimation, in part, because these notions are strongly conflated in the English language. Would the same confusion be found in speakers of other languages? It was predicted that speakers of 'Distance Languages' (i.e., English and Indonesian) would show a strong effect of distance on time estimation when performing the growing line task, whereas speakers of 'Quantity Languages' (i.e., Spanish and Greek) would show a weaker effect.

A complementary 'quantity interference' task was developed, in which participants watched a schematically drawn container of water filling up, one row of pixels at a time, and estimated either how full it became or how much time it remained on the computer screen, using mouse clicks. Behavioral predictions for the Filling Tank task were the mirror image of predictions for the Growing Line task: speakers of Quantity Languages should show a considerable influence of 'fullness' on time estimation, whereas speakers of Distance Languages should show a milder effect.

Participants

A total of 179 subjects participated in Experiments 9 and 10 in exchange for payment. Of these, 53 participants were removed from the analyses reported here for performing the experiment incorrectly or for excessively poor performance, according to the criteria explained in section 2.1. Native English and Spanish speaking participants were recruited from the MIT community, and were tested on MIT campus. Native Indonesian speakers were recruited from the Jakarta community, and were tested at the Cognation Outpost in the Jakarta Field Station of the Max Planck Center for Evolutionary Anthropology. Native Greek speakers were recruited from the Aristotle University of Thessaloniki community, and tested at the University.

Materials and Procedure

Materials and procedures for Experiment 9 (Growing Lines) were identical to those described for Experiment 7a, with the following exception. The instructions for the Indonesian, Greek, and Spanish speaking participants were translated by native speakers of these languages. For all versions of the instructions, care was taken to avoid using any spatial metaphors for time. Although the language of the instructions differed across language groups, the task itself comprised only nonlinguistic stimuli and responses, which were identical for all groups.

The materials and procedures for Experiment 10 (Filling Tanks) were closely analogous to those for Experiment 9. Rather than viewing growing lines, participants viewed 162 containers, and were asked to imagine that each was a tank filling with water. Containers were simple line drawings, 600 pixels high and 500 pixels wide. Empty containers filled gradually, one row of pixels at a time, for varying durations and 'volumes,' and they disappeared when they reached their maximum fullness. Nine durations were fully crossed with nine volumes to produce 81 distinct trial types. Durations ranged from 1000 milliseconds to 5000 milliseconds in 500 millisecond increments. Water levels ranged from 100 to 500 pixels, in 50 pixel increments. As in Experiment 9, however, only the middle five durations and displacements were analyzed as target trials. The two highest and lowest durations and displacements were treated as filler trials, for the reasons described in section 1.7. For each trial, participants estimated either the amount of water in the container (by clicking the mouse once at the bottom of the container and a second time at the appropriate 'water level'), or they estimated the amount of time that the container took to fill (by clicking the hourglass icon, waiting the appropriate time, and clicking it again, as the Growing Line experiments). As before, written instructions were given prior to the start of the task in the native language of the participant. Care was taken to avoid using distance or quantity metaphors for time in the instructions. The task itself was entirely non-linguistic, consisting of filling containers (stimuli) and mouse clicks (responses).

Results and Discussion

Results of the Experiment 9 (Growing Lines) for Indonesian, Greek, and Spanish speakers are summarized in figures 14a-d, 15a-d, and 16a-d, respectively. For English speakers' results on this task, see figure 8a-d. Results of the Experiment 10 (Filling Tanks) for English, Indonesian, Greek, and Spanish speakers are summarized in figures 17a-d, 18a-d, and 19a-d, and 20a-d, respectively. The effects of spatial interference on duration estimation followed predictions based on spatiotemporal metaphors in English, Indonesian, Greek, and Spanish. English and Indonesian speakers showed a strong effect of distance but a weak effect of quantity on time estimation; Greek and Spanish speakers showed the opposite pattern of results (figure 21). A 4 x 2 factorial ANOVA compared the slopes of the effects of target distance and target fullness on time estimation, with Language (English, Indonesian, Greek, Spanish) and Task (Growing Lines, Filling Tanks) as between-subject factors. Results showed a highly significant Language by Task interaction ($F(3,126) = 4.82$, $p<0.003$), with no main effects, signaling a true crossover interaction.

The observed differences in the effects of distance and quantity on duration estimation cannot be attributed to overall differences in performance across tasks or across groups. Within-domain performance (i.e., the effect of target duration on estimated duration; the effect of target distance or fullness on estimated distance or fullness) was compared across tasks and across groups: no significant differences were found between correlations or slopes, even in pairwise comparisons. Furthermore, although the effect of spatial interference on time estimation differed dramatically across groups and tasks, as predicted by the *Deep View* of language-thought relations, the effect of temporal interference on space estimation did not differ. A 4 x 2 factorial ANOVA compared the slopes of the effects of target duration on estimated distance and estimated fullness, with Language (English, Indonesian, Greek, Spanish) and Task (Growing Lines, Filling Tanks) as between-subject factors. Results revealed no significant main effects or interactions[10].

Two further analyses were conducted to explore the relation between language and performance on the time estimation tasks. First, data from speakers of Distance Languages

---

[10] One difference between the Growing Line and Filling Tank tasks was that the lines grew horizontally, but the tanks filled vertically. To rule out the spatial orientation of the stimuli and responses as a potential source of the observed cross-linguistic differences in performance on the Growing Lines and Filling Tank tasks, the Upward Growing Lines task (Experiment 7c) was administered to speakers of Indonesian, Greek, and Spanish. No significant difference was found in the effect of vertical displacement on time estimation across languages. Results suggested that the orientation of stimuli cannot account for the between-group differences observed in Experiments 9 and 10 (see Appendix 2 for summary of Upward Growing Line experiments).

(English, Indonesian) were pooled, and compared with pooled data from speakers of Quantity Languages (Greek, Spanish). This analysis was important for distinguishing effects of language from more general effects of geography or culture. Based on the metaphors in their native languages, participants who presumably had different educational and socioeconomic profiles (e.g. English speakers from the MIT community and Indonesian speakers form the general population in Jakarta) were placed in the same group, while participants whose educational and socioeconomic profiles were presumed to be relatively similar (e.g., native English and native Spanish speaking MIT students and their spouses) were placed in separate groups. A 2 x 2 ANOVA compared the slopes of the effects of target distance and target fullness on time estimation, with Language (Distance Language, Quantity Language) and Task (Growing Lines, Filling Tanks) as between-subject factors. As in the analysis of individual languages, results showed a highly significant Language by Task interaction ($F(1,126) = 13.61$, $p<0.001$), with no main effects (figure 22), consistent with predictions based on the *Deep View* of language-thought relations. Post-hoc independent sample t-tests between groups revealed that the effect of distance interference on time estimation was greater in Distance Language speakers than in Quantity Language speakers (difference of slopes = 1.34, $t(64) = 3.57$, $p<.001$), whereas the effect of quantity interference on time estimation was greater in Quantity Language speakers than in Distance Language speakers (difference of slopes = 0.62, $t(58) = 1.66$, $p<.05$). Post-hoc independent sample t-tests between tasks revealed that the effect of distance interference on time estimation (Growing Lines) was greater than the effect of quantity interference on time estimation (Filling Tanks) in Distance Language speakers (difference of slopes = 1.07, $t(60) = 2.73$, $p<.004$). By contrast, the effect of quantity interference on time estimation (Filling Tanks) was greater than the effect of distance interference on time estimation (Growing Lines) in Quantity Language speakers (difference of slopes = 0.88, $t(62) = 2.47$, $p<.008$).

Finally, an analysis was conducted to quantify the relation between linguistic metaphors and performance on the nonlinguistic time estimation tasks. Using the corpus data reported in Experiment 8, an asymmetry ratio (AR) was computed in order to express the relative prevalence of distance and quantity metaphors in each language as a value on a scale from –1 to 1:

$$\text{AR\_metaphors} = \frac{(\text{proportion distance metaphors} - \text{proportion quantity metaphors})}{(\text{proportion distance metaphors} + \text{proportion quantity metaphors})}$$

A positive AR_metaphors indicated a preference for distance metaphors and a negative AR_metaphors indicated a preference for quantity metaphors, according to the corpus data.

Likewise, an asymmetry ratio was computed in order to express the relative effects of distance and quantity interference on time estimation for speakers of each language as a value on a scale from –1 to 1:

$$\text{AR\_slopes} = \frac{(\text{effect of distance on time estimation} - \text{effect of quantity on time estimation})}{(\text{effect of distance on time estimation} + \text{effect of quantity on time estimation})}$$

A positive AR_slopes indicated a greater effect of distance interference on time estimation, and a negative AR_slopes indicated a greater effect of quantity interference on time estimation, according to the data from the Growing Line and Filling Tank tasks (Experiments 9 and 10).

The asymmetry of slopes for speakers of English, Indonesian, Spanish, and Greek was plotted as a function of the asymmetry of metaphors in these languages, and a nonparametric correlation was computed. Results showed a perfect rank order correlation (Kendall's Tau_b = 1.00, $p<.02$), demonstrating a strong association between linguistic metaphors for duration and nonlinguistic mental representations of time.

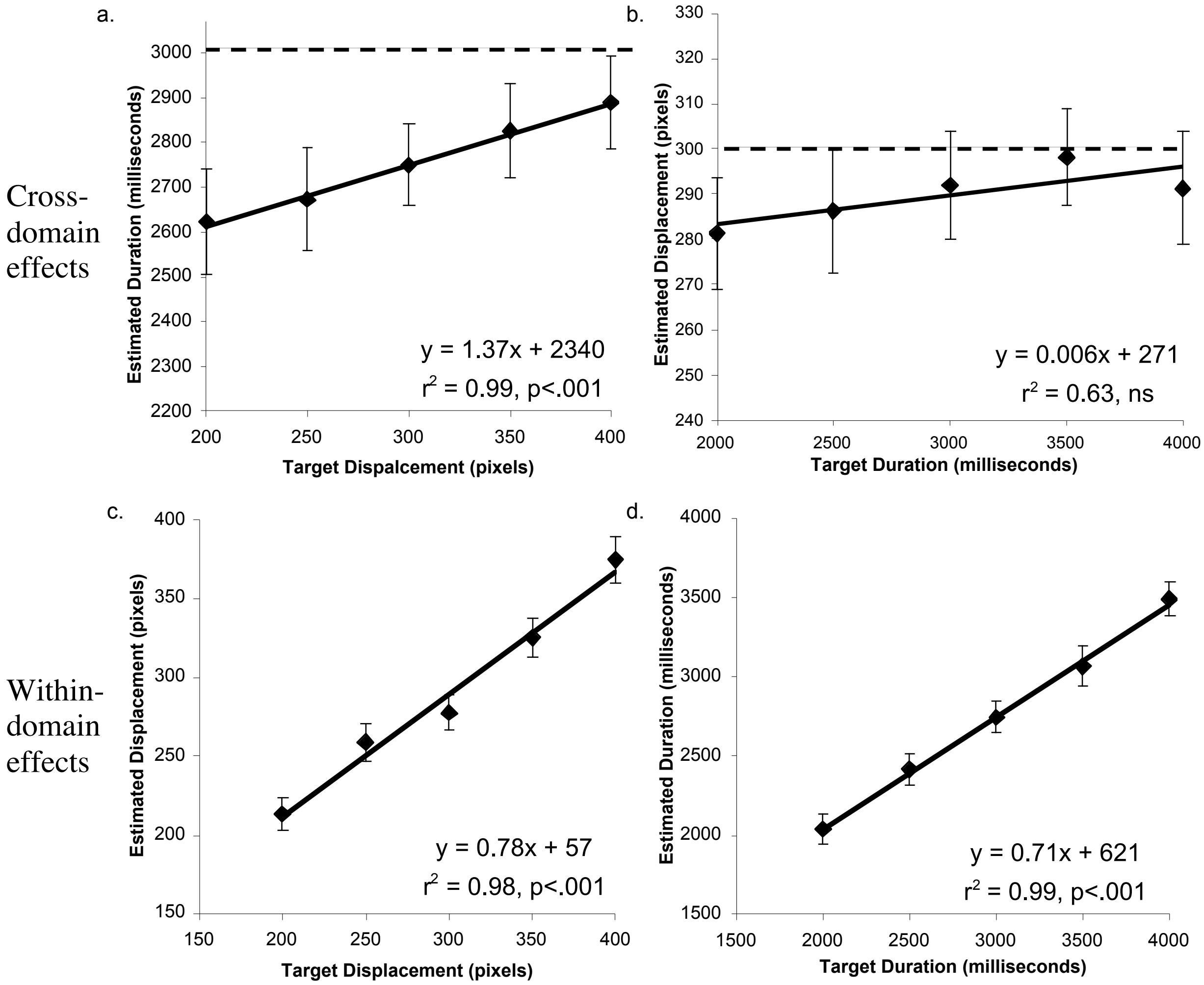

Figure 14. Grand averaged duration and displacement estimates (n=17) for Experiment 9b (Growing Lines, Indonesian Speakers). Top: Cross-domain effects. 14a. (left) Effect of displacement on duration estimation. 14b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 14a. and 14b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 14c. (left) Effect of target displacement on estimated displacement. 14d. (right) Effect of target duration on estimated duration.

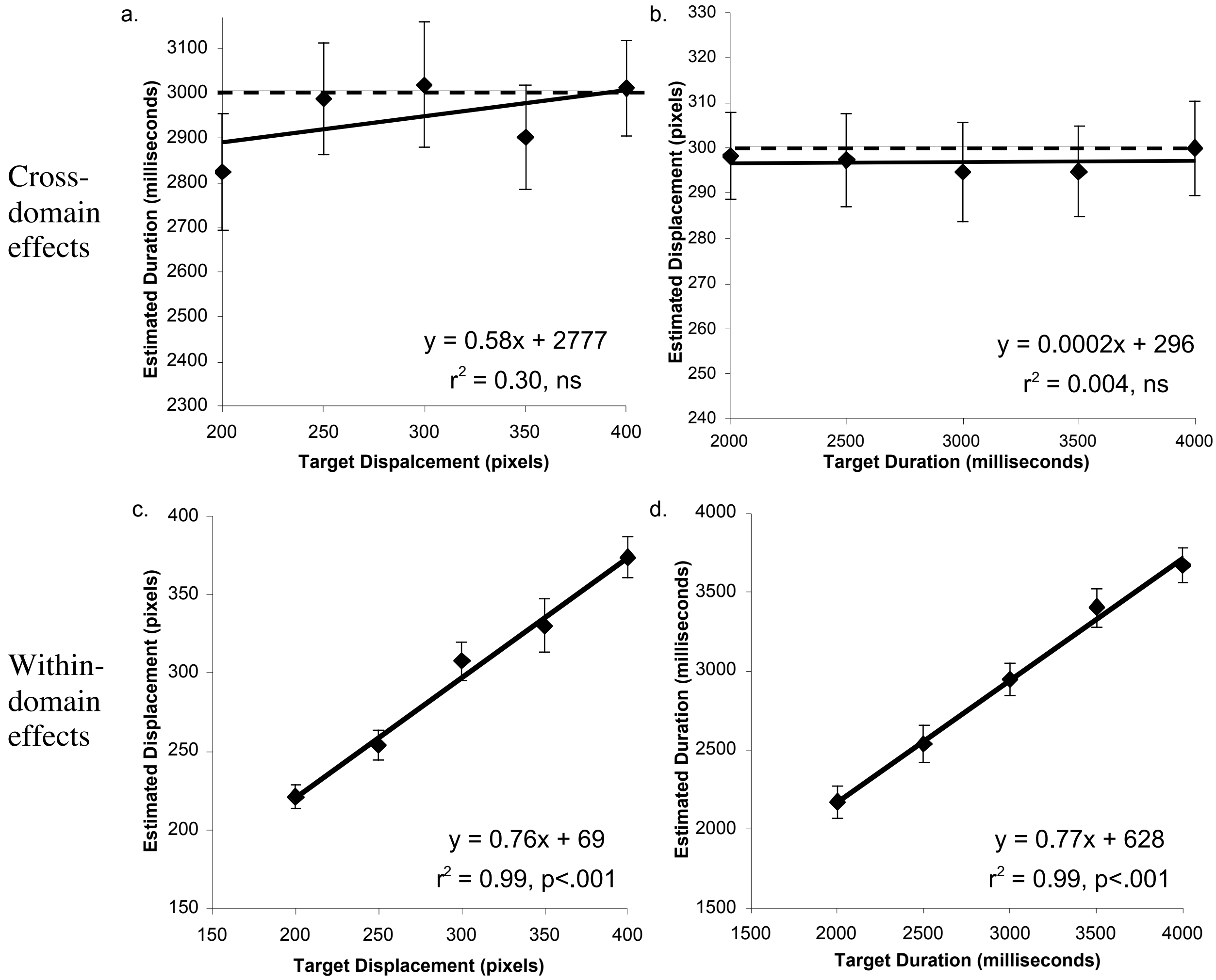

Figure 15. Grand averaged duration and displacement estimates (n=14) for Experiment 9c (Growing Lines, Greek Speakers). Top: Cross-domain effects. 15a. (left) Effect of displacement on duration estimation. 15b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 15a. and 15b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 15c. (left) Effect of target displacement on estimated displacement. 15d. (right) Effect of target duration on estimated duration.

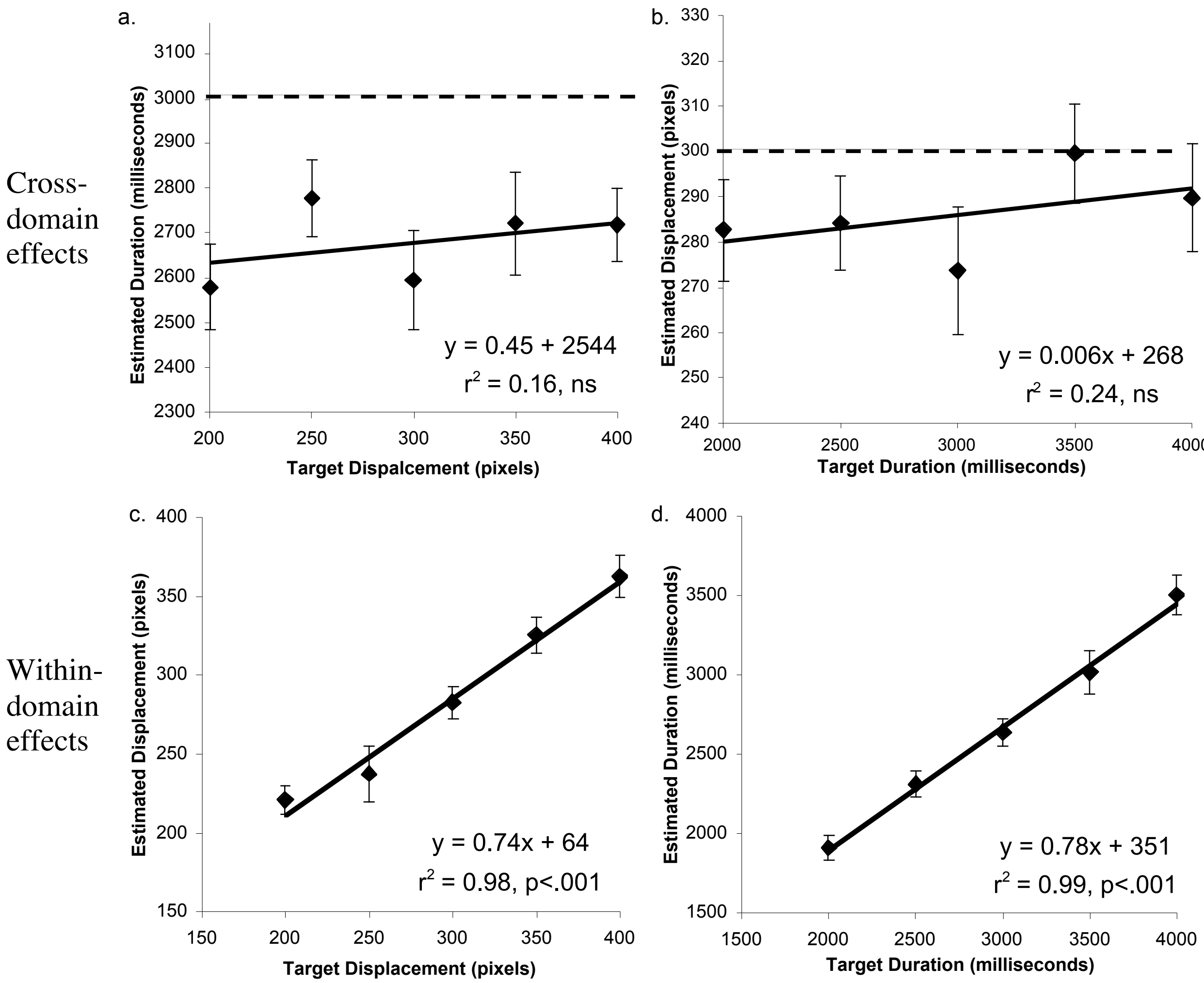

Figure 16. Grand averaged duration and displacement estimates (n=14) for Experiment 9d (Growing Lines, Spanish Speakers). Top: Cross-domain effects. 16a. (left) Effect of displacement on duration estimation. 16b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 16a. and 16b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 16c. (left) Effect of target displacement on estimated displacement. 16d. (right) Effect of target duration on estimated duration.

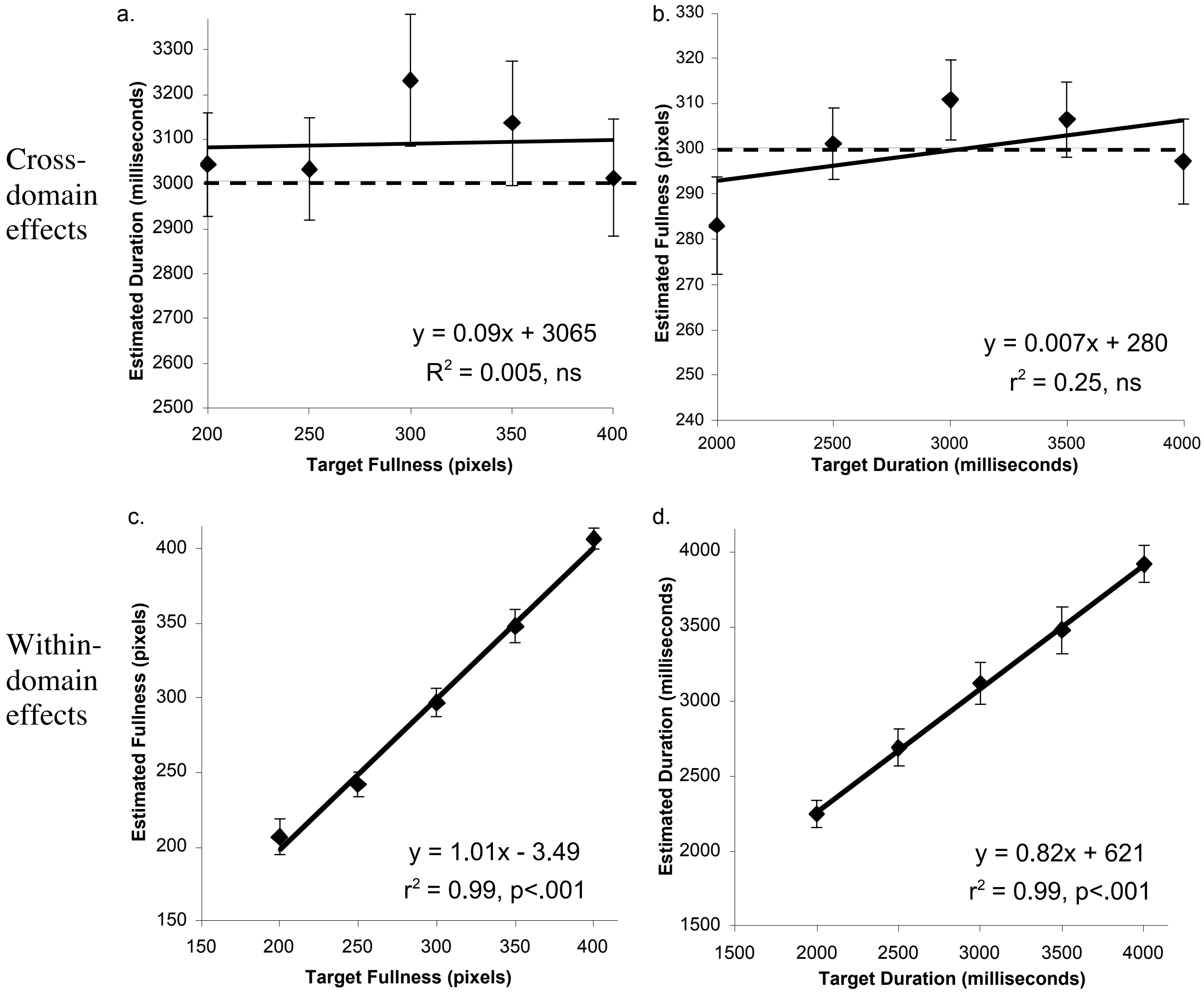

Figure 17. Grand averaged duration and fullness estimates (n=16) for Experiment 10a (Filling Tanks, English Speakers). Top: Cross-domain effects. 17a. (left) Effect of fullness on duration estimation. 17b. (right) Effect of duration on fullness estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 17a. and 17b. are proportionate with respect to the total range of target duration and fullness. Bottom: Within-domain effects. 17c. (left) Effect of target fullness on estimated fullness. 17d. (right) Effect of target duration on estimated duration.

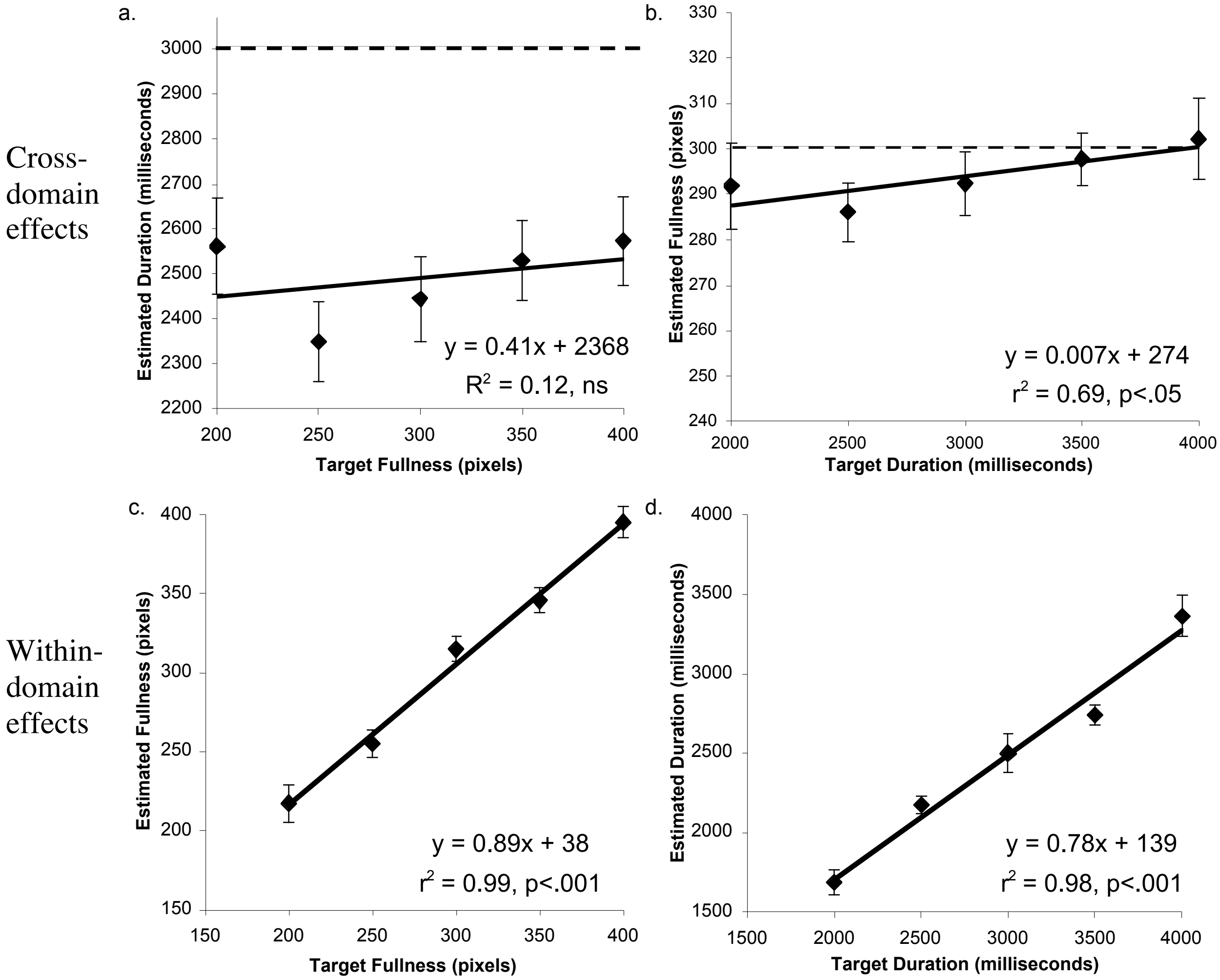

Figure 18. Grand averaged duration and fullness estimates (n=16) for Experiment 10b (Filling Tanks, Indonesian Speakers). Top: Cross-domain effects. 18a. (left) Effect of fullness on duration estimation. 18b. (right) Effect of duration on fullness estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 18a. and 18b. are proportionate with respect to the total range of target duration and fullness. Bottom: Within-domain effects. 18c. (left) Effect of target fullness on estimated fullness. 18d. (right) Effect of target duration on estimated duration.

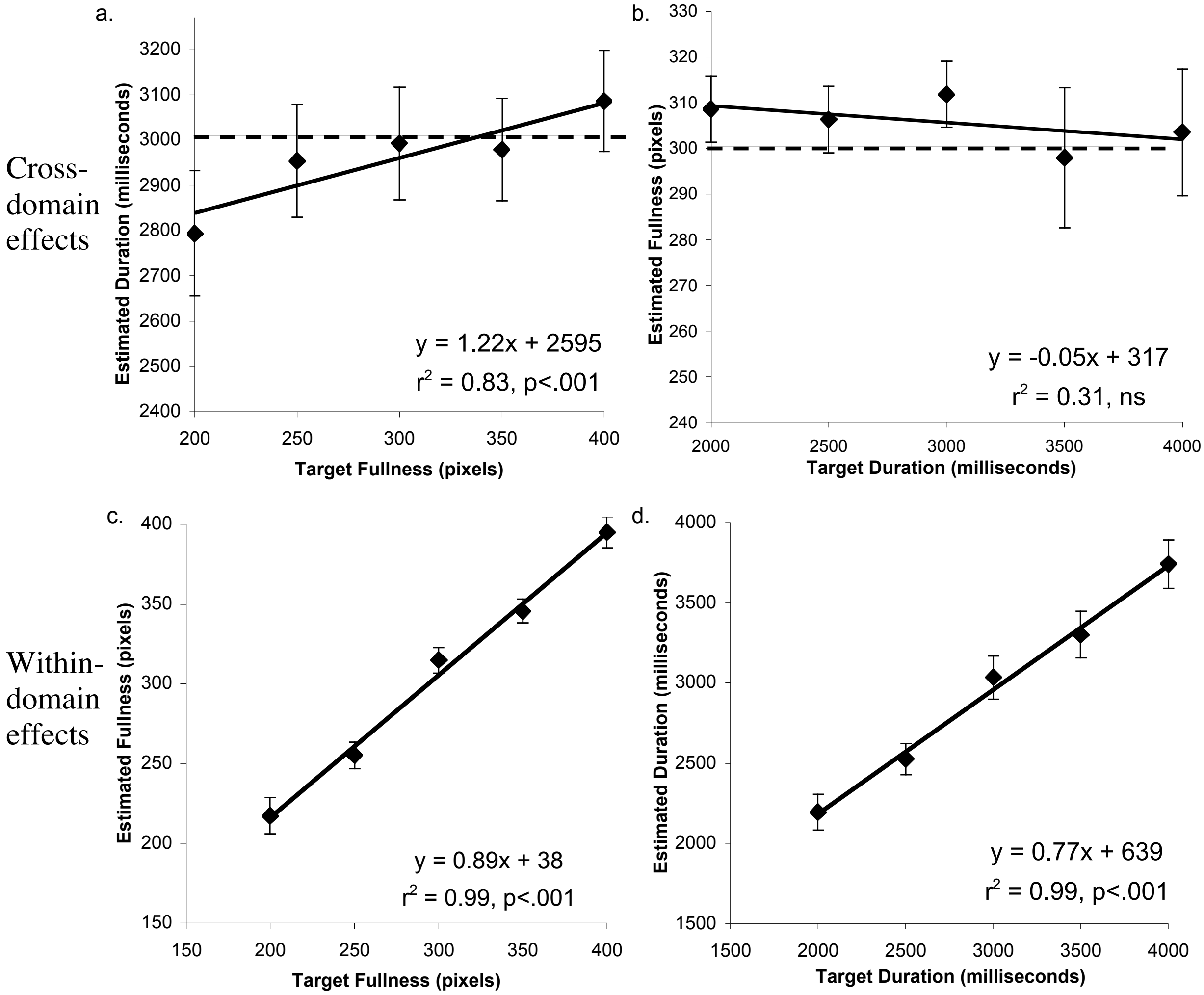

Figure 19. Grand averaged duration and fullness estimates (n=14) for Experiment 10c (Filling Tanks, Greek Speakers). Top: Cross-domain effects. 19a. (left) Effect of fullness on duration estimation. 19b. (right) Effect of duration on fullness estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 19a. and 19b. are proportionate with respect to the total range of target duration and fullness. Bottom: Within-domain effects. 19c. (left) Effect of target fullness on estimated fullness. 19d. (right) Effect of target duration on estimated duration.

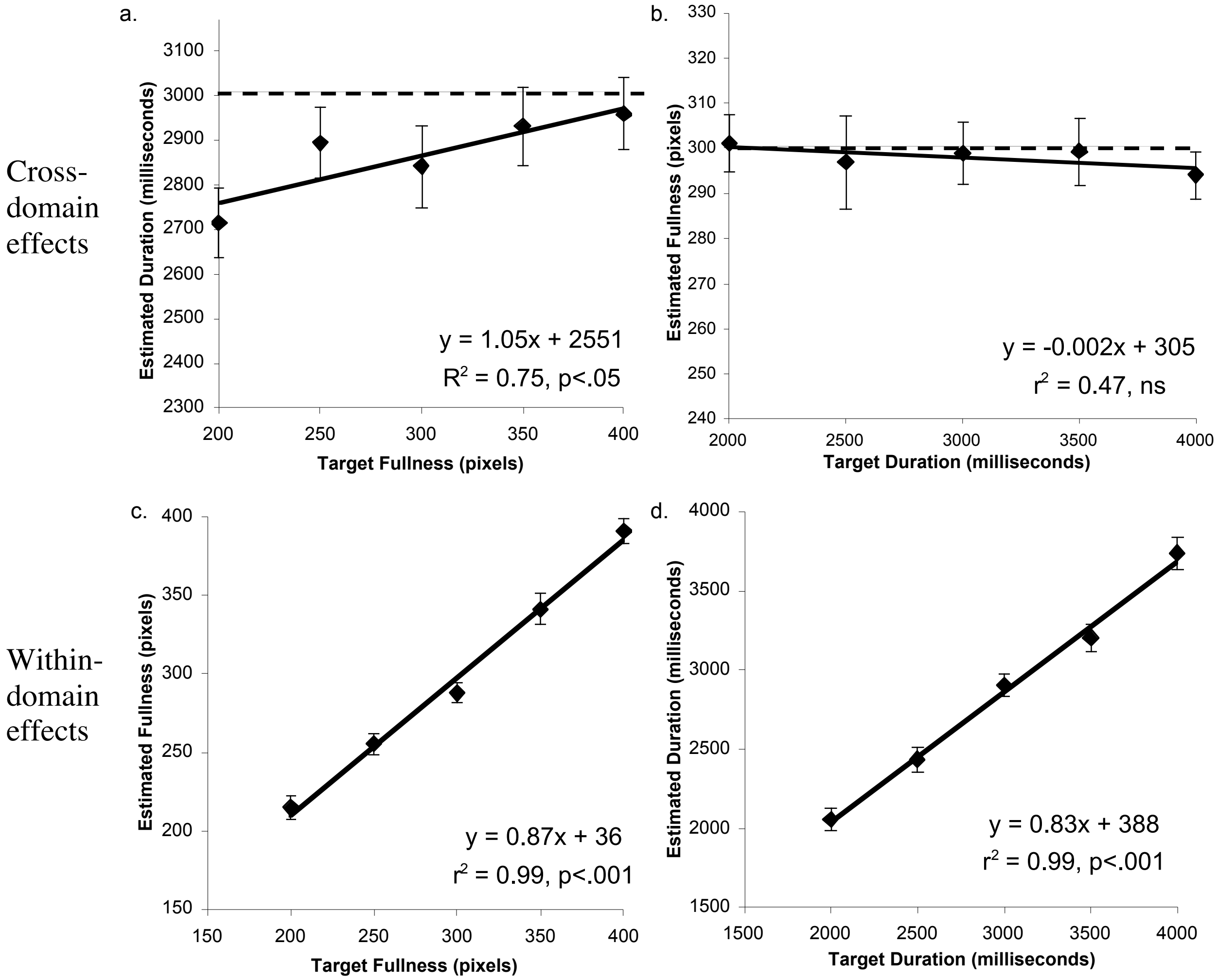

Figure 20. Grand averaged duration and fullness estimates (n=18) for Experiment 10d (Filling Tanks, Spanish Speakers). Top: Cross-domain effects. 20a. (left) Effect of fullness on duration estimation. 20b. (right) Effect of duration on fullness estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 20a. and 20b. are proportionate with respect to the total range of target duration and fullness. Bottom: Within-domain effects. 20c. (left) Effect of target fullness on estimated fullness. 20d. (right) Effect of target duration on estimated duration.

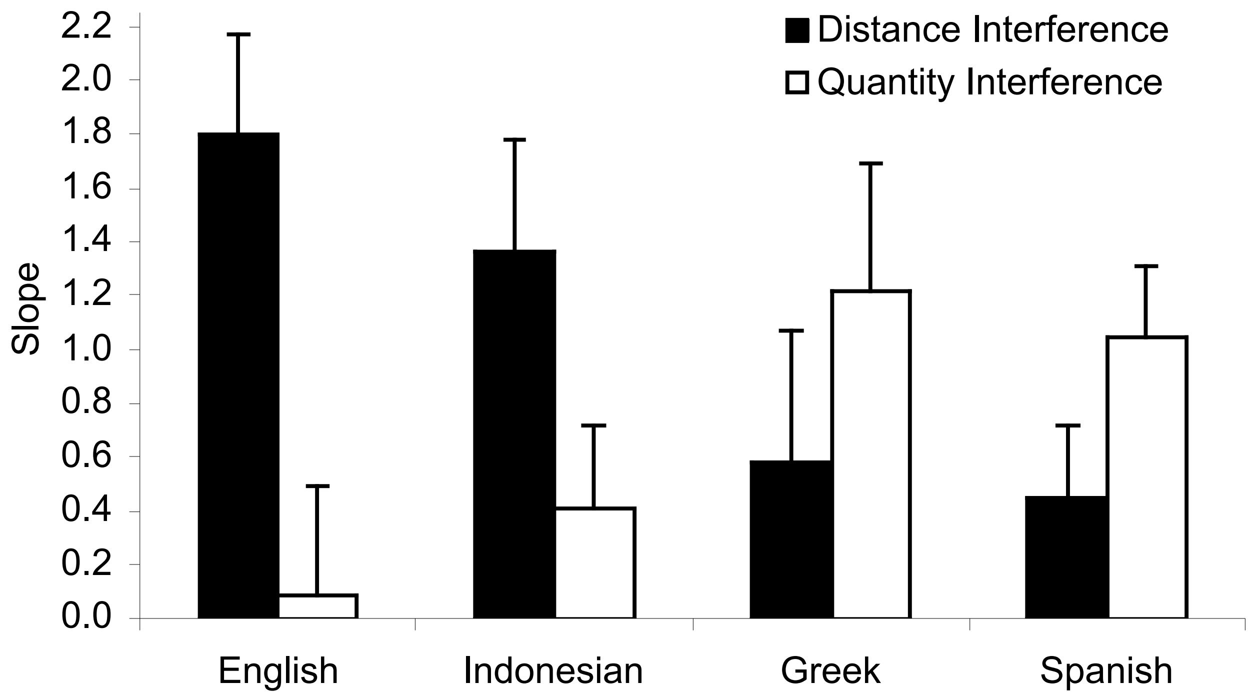

Figure 21. Summary of the effects of distance interference (black bars) and quantity interference (white bars) on duration estimation for Experiments 9a-d (Growing Lines) and 10a-d (Filling Tanks).

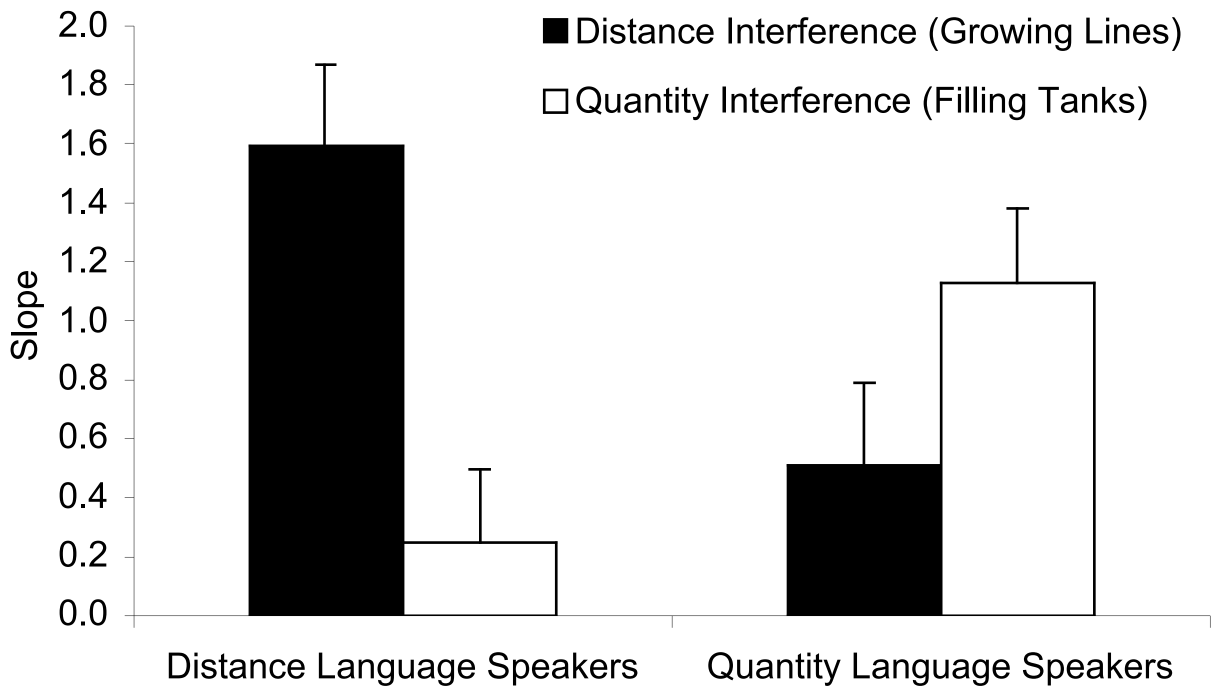

Figure 22. Summary of the effects of distance interference (black bars) and quantity interference (white bars) on duration estimation in combined data for speakers of Distance Languages (English, Indonesian) and Quantity Languages (Greek, Spanish).

3.3 How does language shape the way we think?

How do people come to think about time in terms of space, and how do speakers of different languages come to represent time differently? One possibility is that people track the kinds of correlations in experience that are important for perceiving and acting on their environment, and they learn associations between time and space by interacting with the physical world (e.g., by observing moving objects and changing quantities)[11]. Since presumably the laws of physics are the same in all language communities, pre-linguistic children's conceptual mappings between time, distance, and quantity could be the same universally. Later, as children acquire language, these mappings could be adjusted: each time we use a linguistic metaphor, we may invoke the corresponding conceptual mapping. Speakers of Distance Languages then would invoke the time-distance mapping frequently, eventually strengthening it at the expense of the time-quantity mapping (and vice-versa for speakers of Quantity Languages).

Did language experience give rise to the language-related differences in performance reported for the Growing Line and Filling Tank experiments? A perennial complaint about studies that purport to show effects of language on thought is that researchers mistake correlation for causation. Although it is difficult to imagine what nonlinguistic cultural or environmental factors could have caused performance on Experiments 9 and 10 in English, Indonesian, Greek, and Spanish speakers to align so uncannily with the metaphors in these languages, the data are nevertheless correlational.

[11] This proposal entails that time can be represented *qua* time, as is explicit in Conceptual Metaphor Theory (Lakoff & Johnson, 1980, 1999) and in the Metaphoric Structuring view (Boroditsky, 2000). Yet, metaphoric representations of time are not obviated if primitive temporal notions are assumed to be too vague, fleeting, or resistant to verbal or imagistic representation to support higher order reasoning about time -- from conscious recollection of temporal intervals to imagining time-travel.

Rigorously testing for a causal influence of language on thought is a hard problem, since experimenters cannot randomly assign subjects to have one first language or another.

For Experiment 11, a pair of training tasks was conducted (a) to provide an *in principle* demonstration that language can influence even the kinds of low-level mental representations that people construct while performing psychophysical tasks, and (b) to test the hypothesis that language shapes time representations in natural settings by adjusting the strengths of cross-domain mappings. During the training task, native English speaking participants compared the duration of events using either distance metaphors or quantity metaphors. Immediately after training, they performed the Filling Tank task, described in Experiment 10. It was predicted that if using a linguistic metaphor invokes the corresponding conceptual mapping between source and target domains (e.g., space and time), then repeatedly using quantity metaphors during training should transiently strengthen participants' nonlinguistic quantity-time mapping, and should increase the effect of quantity on time estimation in the Filling Tank task. Repeatedly using distance metaphors during training should weaken participants' nonlinguistic quantity-time mapping, and should decrease the effect of quantity on time estimation in the Filling Tank task.

Participants

A total of 36 subjects from the Stanford University community participated in Experiments 11a and 11b, in exchange for payment. All were native monolingual English speakers, according to a language background questionnaire (i.e., English was the only language they learned before age 5, and was their strongest language at time of test). Of these, 6 participants were removed from the analyses reported here for performing the Filling Tanks experiment incorrectly or for excessively poor performance, according to the criteria explained in section 2.1.

Materials and Procedure

Participants were randomly assigned to perform either a Distance Training or Quantity Training task. Participants completed 192 fill-in-the-blank sentences using the words *longer* or *shorter* for Distance Training, and *more* or *less* for the Quantity Training task. Half of the sentences compared the length or capacity of physical objects (e.g., An alley is *longer* / *shorter* than a clothesline; A teaspoon is *more* / *less* than an ocean), the other half compared the duration of events (e.g., A sneeze is *longer* / *shorter* than a vacation; A sneeze is *more* / *less* than a vacation). By using distance terms to compare event duration, English speakers were reinforcing the already preferred source-target mapping between distance and time. By using quantity terms, English speakers were describing event durations similarly to speakers of a Quantity Language (see Greek examples in section 2.2), and by hypothesis, they were invoking the dispreferred quantity-time mapping. After training, all participants performed the Filling Tank task used in Experiment 10.

Results and Discussion

Results of the training phase showed that participants filled in the blanks with high accuracy for both the Distance Training task ($\underline{M}_{\%\ \text{correct}}$ = 0.97, SD = 0.02) and the Quantity Training task ($\underline{M}_{\%\ \text{correct}}$ = 0.98, SD = 0.02). One-sample t-tests showed that accuracy was significantly above chance (50%) for both tasks, and an independent samples t-test showed that accuracy did not differ between tasks (difference of means = 0.01, $t(28) = 1.10$, ns).

Results of the post-Distance Training Filling Tanks task (Experiment 11a) are summarized in figure 23a-d, and results of the post-Quantity Training Filling Tanks task (Experiment 11b) are summarized in figure 24a-d. Consistent with predictions, the effect of quantity on time estimation was nonsignificant following Distance Training (as was also the case in untrained English participants), but was significant following Quantity Training. Compared with data from the untrained English speakers reported in Experiment 10, the slope of the effect quantity on time estimation of was increased following Quantity Training and decreased following Distance Training, though these differences from untrained participants were not statistically significant. Importantly, however, the slope of the effect of quantity on time estimation (figure 25) was significantly greater after Quantity training than after Distance training (difference of slopes = 0.89, $t(28) = 1.73$, $p<.05$). Whereas the slope of the effect of quantity on time estimation in untrained English speakers was significantly lower than in untrained Greek speakers (difference of slopes = 1.13, $t(28) = 1.84$, $p<.04$) and untrained Spanish speakers (difference of slopes = 0.97, $t(32) = 2.06$, $P<.02$) reported in Experiment 10, no significant difference was found when the slope of the effect of quantity on time estimation in Quantity Trained English speakers was compared with the slopes in untrained Greek or Spanish speakers. Linguistic training with quantity metaphors caused native English speakers to perform the Filling Tank task more like native speakers of Greek or Spanish, while linguistic training with distance metaphors had the opposite effect.

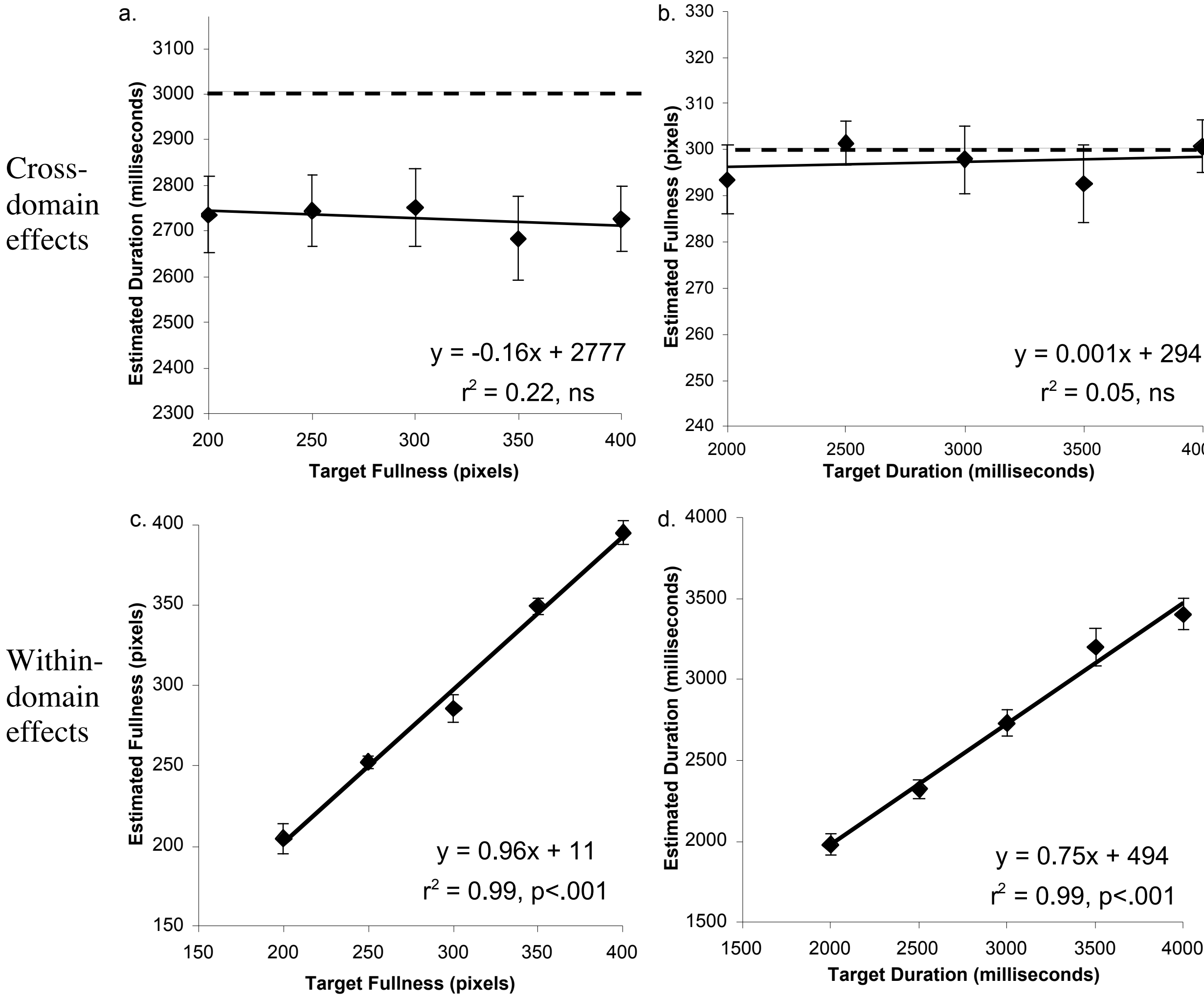

Figure 23. Grand averaged duration and fullness estimates (n=15) for Experiment 11a (Filling Tanks, Distance-Trained English Speakers). Top: Cross-domain effects. 23a. (left) Effect of fullness on duration estimation. 23b. (right) Effect of duration on fullness estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 23a. and 23b. are proportionate with respect to the total range of target duration and fullness. Bottom: Within-domain effects. 23c. (left) Effect of target fullness on estimated fullness. 23d. (right) Effect of target duration on estimated duration.

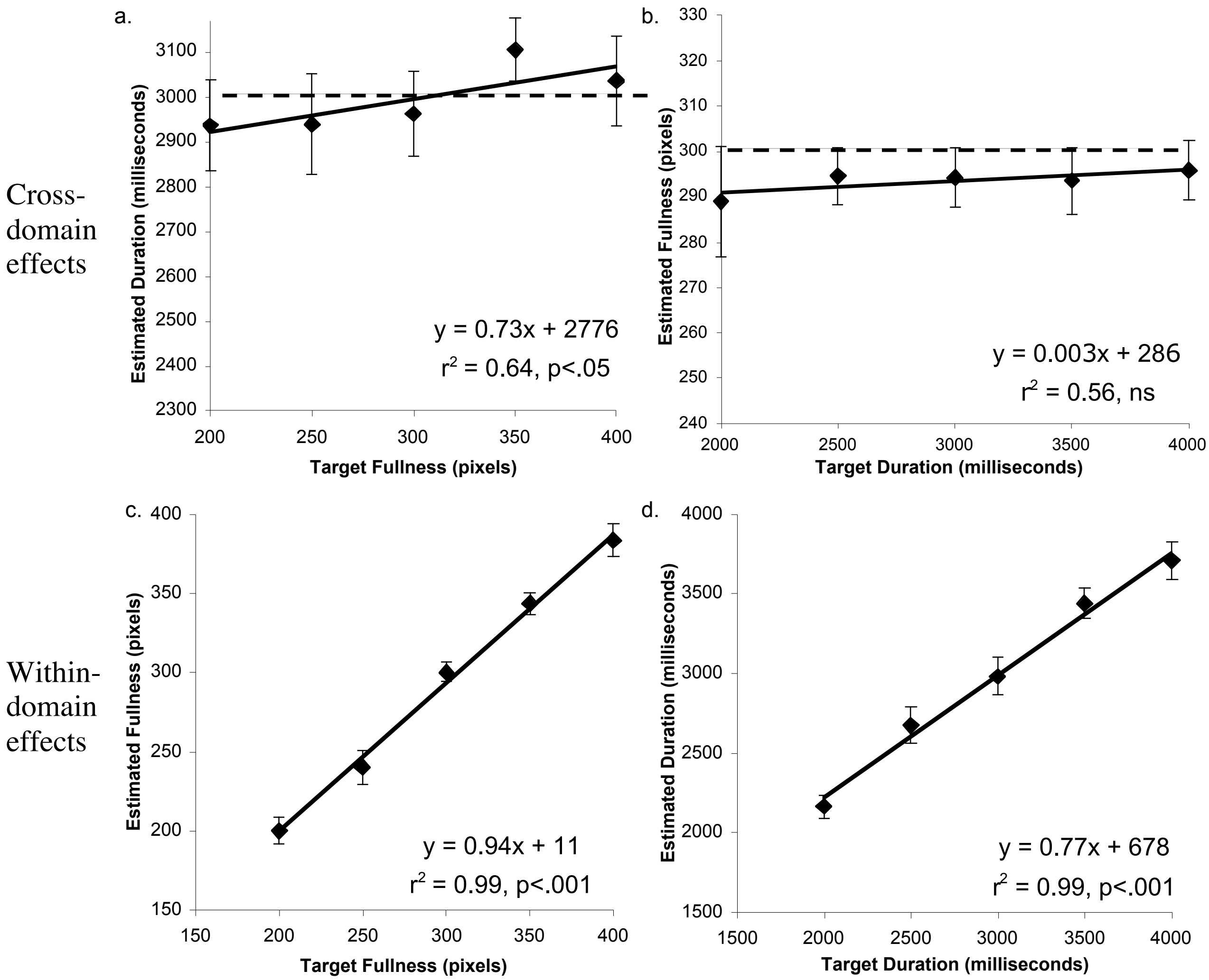

Figure 24. Grand averaged duration and fullness estimates (n=15) for Experiment 11b (Filling Tanks, Quantity-Trained English Speakers). Top: Cross-domain effects. 24a. (left) Effect of fullness on duration estimation. 24b. (right) Effect of duration on fullness estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 24a. and 24b. are proportionate with respect to the total range of target duration and fullness. Bottom: Within-domain effects. 24c. (left) Effect of target fullness on estimated fullness. 24d. (right) Effect of target duration on estimated duration.

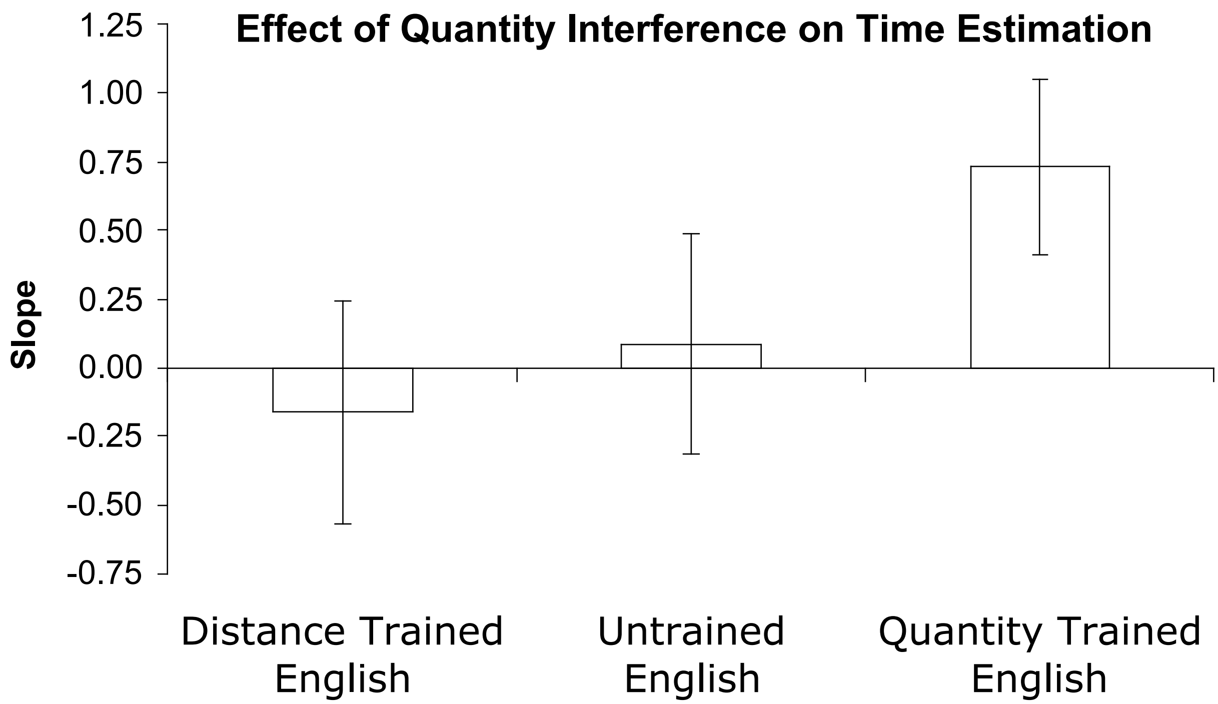

Figure 25. Effects of quantity interference on duration estimation in Distance-Trained English speakers (Experiment 11a, left) and Quantity-Trained English speakers (Experiment 11b, right). Data from untrained English Speakers (middle) reported in Experiment 10a are included as a point of reference.

Chapter 4: Conclusions

Together, the experiments presented here constitute a body of evidence that is inconsistent with the *Shallow View* of language-thought relations, and provide some of the first evidence for the view that language has *Deep* influences on nonlinguistic mental representation. Experiments 1-7 show that people not only talk about time in terms of space, they also think about time using spatial representations -- even when they're performing nonlinguistic tasks. Experiments 8-10 show that people who talk differently about time also think about it differently, in ways that correspond to their language-particular metaphors. Experiment 11 shows that language not only reflects the structure of underlying mental representations of time, it can also shape those representations in ways that influence how people perform even low-level, nonlinguistic, perceptuo-motor tasks.

These findings are difficult to reconcile with a 'universalist' position according to which language calls upon nonlinguistic concepts that are presumed to be "universal" (Pinker, 1994, pg. 82) and "immutable" (Papafragou, Massey, & Gleitman, 2002, pg. 216). Beyond influencing "thinking for speaking" (Slobin, 1986), our particular languages can also influence the nonlinguistic representations we build for remembering, acting on, and perhaps even perceiving the world around us. It may be universal that people map time onto space according to the metaphors in their language, but as a consequence, members of different language communities develop distinctive conceptual repertoires.

Direct evidence that spatial cognition supported the evolution of temporal cognition may forever elude us, because human history cannot be recreated in the laboratory, and the mind leaves no fossil record. However, the studies reported here demonstrate the primacy of spatial representations in the mind that evolution produced. Previously, inferences about the perceptual foundations of abstract thinking have rested principally on linguistic data (and, recently, psycholinguistic data). These psychophysical experiments show that even low-level perceptuo-motor representations in the domains of space and time are related just as predicted by linguistic metaphors, and just as expected if the more abstract domain arose as an exaptation from the more concrete. The structure of abstract domains like time appears to depend, in part, on both linguistic experience and on physical experience in perception and motor action.

References

Boroditsky, L. (2000). Metaphoric structuring: understanding time through spatial metaphors. *Cognition, 75*(1), 1-28.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43*(1), 1-22.

Boroditsky, L., & Ramscar, M. (2002). The Roles of Body and Mind in Abstract Thought. *Psychological Science, 13*(2), 185-189.

Casasanto, D., & Boroditsky, L. (2003). *Do we think about time in terms of space?* Paper presented at the 25th Annual Conference of the Cognitive Science Society, Boston.

Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, S., et al. (2004). *How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish*. Paper presented at the 26th Annual Conference of the Cognitive Science Society, Chicago.

Chomsky, N. (1975). *Reflections on Language*. New York: Norton & Company.

Clark, E. (2003). Languages and Representations. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind* (pp. 17-23). Cambridge: MIT Press.

Clark, H. H. (1973). Space, Time, Semantics and the Child. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language* (pp. 27-63). New York: Academic Press.

Clark, H., & Brownell, H. (1976). Position, direction, and their perceptual integrality. *Perception & Psychophysics, 19*(4), 328-334.

Cohen, J. (1967). *Psychological time in health and disease*. Springfield: Charles C. Thomas.

Darwin, C. (1859/1998). *The Origin of Species By Means of Natural Selection or The Preservation of Favored Races in the Struggle for Life*. New York: The Modern Library.

Darwin, C. (1874/1998). *The Descent of Man.* Amherst: Promethius Books.

DeLong, A. (1981). Phenomenological Space-Time: toward an Experiential Relativity. *Science, 213*(4508), 681-683.

Fodor, J. (1975). *The Language of Thought*. Cambridge: Harvard University Press.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Gentner, D. (Ed.). (2001). *Spatial Metaphors in Temporal Reasoning*. Cambridge: MIT Press.

Gibbs, R. W., jr. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge: Cambridge University Press.

Glicksohn, J. (2001). Temporal cognition and the phenomenology of time: a multiplicative function for apparent duration. *Consciousness and Cognition, 10*, 1-25.

Gould, S., & Vrba, E. (1982). Exaptation -- a missing term in the Science of Form. *Paleobiology, 8*, 4-15.

Gould, S. J. (1980). Natural Selection and the Brain: Darwin vs. Wallace. In *The Panda's Thumb* (pp. 47-58). New York: Norton.

Gould, S. J. (1991). Not Necessarily a Wing. In *Bully for Bronntosaurus* (pp. 139-151). New York: Norton & Co.

Gruber, J. (1965). *Studies in Lexical Relations*. MIT, Cambridge.

Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.

Jackendoff, R. (1983). *Semantics and Cognition*. Cambridge: MIT PRess.

James, W. (1890). *Principles of psychology*. New York: Rinehart & Winston.

Katz, J., & Fodor, J. (1963). The Structure of a Semantic Theory. *Language, 39*(2), 170-210.

Lafargue, P. (1898). Ursprung der abstrakten Ideen. *Die Neue Zeit, XVIII*(2).

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Chicago: University of Chicago Press.

Langacker, R. (1987). An introduction to Cognitive Grammar. *Cognitive Science, 10*, 1-40.

Levelt, W. (1989). *Speaking: from intention to articulation*. Cambridge: MIT Press.

Locke, J. (1689/1995). *An Essay Concerning Human Understanding*. Amherst: Promethius Books.

Mach, E. (1896/1897). *Contributions to the Analysis of Sensations*. Chicago: Open Court Publishing Company.

Murphy, G. (1996). On Metaphoric Representation. *Cognition, 60*, 173-204.

Murphy, G. (1997). Reasons to doubt the present evidence for metaphoric representation. *Cognition, 62*, 99-108.

Narayanan, S. (1997). *Embodiment in Language Understanding: Sensory-Motor Representations for Metaphoric Reasoning About Event Descriptions*. University of California, Berkley, Berkley.

Ornstein, R. (1969). *On the experience of time*. Hammondsworth: Penguin.

Papafragou, A., Massey, C., & Gleitman, L. (2002). Shake, Rattle, 'n 'Roll: the representation of motion in langauge and cognition. *Cognition, 84*, 189-219.

Pinker, S. (1997). *How the mind works*. New York: Norton.

Pullum, G. (1991). *The Great Eskimo Vocabulary Hoax*. Chicago: University of Chicago Press.

Slobin, D. (1986). From "THOUGHT AND LANGUAGE" to "THINKING FOR SPEAKING". In (pp. 71-96).

Sommer, F. (1963). Types of Ontology. *Philosophical Review, 72*, 327-363.

Snedeker, J., & Gleitman, L. (2004). Why is it hard to label our concepts? In Hall & Waxman (Eds.), *Weaving a Lexicon*. Cambridge: MIT Press.

Talmy, L. (1988). Force Dynamics in Language and Cognition. *Cognitive Science, 12*, 49-100.

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology, 23*, 515-557.

Wallace, A. (1870/2003). Contributions to the theory of natural selection. In *Contributions to the theory of natural selection* (pp. 400). London: Routledge.

Whorf, B. L. (1939/2000). Language, thought, and reality: selected writings of Benjamin Lee Whorf. Cambridge: MIT Press.

Zakay, D., & Block, R. (1997). Temporal Cognition. *Current Directions in Psychological Science, 6*(1), 12-16.

Appendix 1

Questionnaire Study: Describing Event Duration with Distance and Quantity Terms

To validate the results of Experiment 8, a questionnaire study was conducted to investigating the use of distance and quantity metaphors for duration across languages. A total of 83 participants (21 native English speakers, 36 native Indonesian speakers, 9 native Greek speakers, and 17 native Spanish speakers) completed a questionnaire. For the English version, twenty-five nouns and ten adjectives selected from the combined Kucera & Francis and Wall Street Journal corpora were fully crossed, to produce 250 adjective-noun pairs. Pairs were composed of the highest frequency common nouns in English that could refer to an event (e.g., day, night, party, war, process), and the highest frequency adjectives in English that could potentially modify the duration of an event in at least one of the languages investigated (e.g., long, short, big, little, much). Participants were asked to indicate whether each noun phrase (e.g., *short party, big night*) described the duration of the event (as opposed to some other property, such as its physical size or importance), and to indicate how likely they would be to use the adjective to modify the duration of the noun. The questionnaire was translated into Indonesian, Greek, and Spanish, and administered to native speakers of the four languages surveyed in Experiment 8.

On the whole, results corroborated those of the Google search comparing the relative frequency of Distance and Quantity metaphors across languages (Experiment 8). Figure 26 shows the proportion of phrases using Distance adjectives (e.g., *long, short*) and the proportion of phrases using Quantity adjectives (e.g., *big, much*) that were judged to modify the duration of an event (as opposed to some other property of the event).

Analyses were conducted (a) to quantify the relation between linguistic metaphors as measured by the Questionnaire Study and by the Google Search, and (b) to quantify the relation between linguistic metaphors as measured by the Questionnaire Study and performance on the nonlinguistic time estimation tasks. An asymmetry ratio (AR) was computed in order to express the relative prevalence of distance and quantity metaphors in each language as measured by the Questionnaire Study on a scale from –1 to 1:

$$\text{AR\_Questionnaire} = \frac{\text{(proportion distance metaphors – proportion quantity metaphors)}}{\text{(proportion distance metaphors + proportion quantity metaphors)}}$$

First, the asymmetry of metaphors as measured by the Google Search (Experiment 8, see section 3.1) was plotted as a function of the asymmetry of metaphors as measured by the Questionnaire Study. Results showed the predicted positive rank order correlation, which was marginally significant (Kendall's Tau_b = 0.67, $p<.09$).

Next, the asymmetry of the slopes of the effects of distance and quantity on time estimation in the Growing Lines and Filling Tanks experiments (Experiments 9 and 10, see section 3.2) was plotted as function of the asymmetry of metaphors as measured by the Questionnaire Study. Again, results showed the predicted positive rank order correlation, which was marginally significant (Kendall's Tau_b = 0.67, $p<.09$).

For English, Indonesian, and Greek speakers, results of the Questionnaire study corresponded remarkably closely to the results of the Google Search, and of the psychophysical time estimation tasks. As expected, questionnaire data showed a preference for Distance metaphors to modify event duration in English and Indonesian speakers, but a preference for Quantity metaphors in Greek speakers. Notably, however, the Questionnaire data for Spanish speakers did not show the expected preference for Quantity metaphors: rather, a sign test evaluating the preference for Distance vs. Quantity metaphors in Spanish speaking participants showed a significant preference for Distance metaphors ($p<.05$). One possible explanation for the discrepancy between the data from the corpus and questionnaire studies is that Spanish may prefer Quantity metaphors for describing time, *per se* (e.g., *mucho tiempo*) but Distance metaphors for describing event duration (e.g., *una larga fiesta*). Alternatively, the questionnaire results may reflect second language ‘contamination’ of native Spanish speaking participants’ judgments. Whereas Indonesian and Greek participants were tested in Indonesia and Greece, respectively, Spanish speaking participants were tested on MIT campus. Most were students, accustomed to speaking, reading, writing, and even teaching in English. It is possible that habitually using English expressions like *a long party* influenced participants’ judgments of Spanish noun phrases like *larga fiesta*.

Also, as mentioned in section 3.1, informal interviews with Spanish speaking participants suggest that distance metaphors for time may be more acceptable in North American dialects of Spanish (possibly due to the influence of English) than in South American and European Spanish dialects. The majority of participants who completed the Questionnaire learned Spanish in North America, and grew up bilingual in English. It is possible that if the study were repeated in monolinguals, or exclusively in South American and European Spanish speakers, that the Questionnaire data would more closely resemble the Google data. (Likewise, results on the psychophysical tasks might show an even stronger difference between the effects of Quantity and Distance on time estimation if Experiments 9 and 10 were run in monolingual Spanish speakers.)

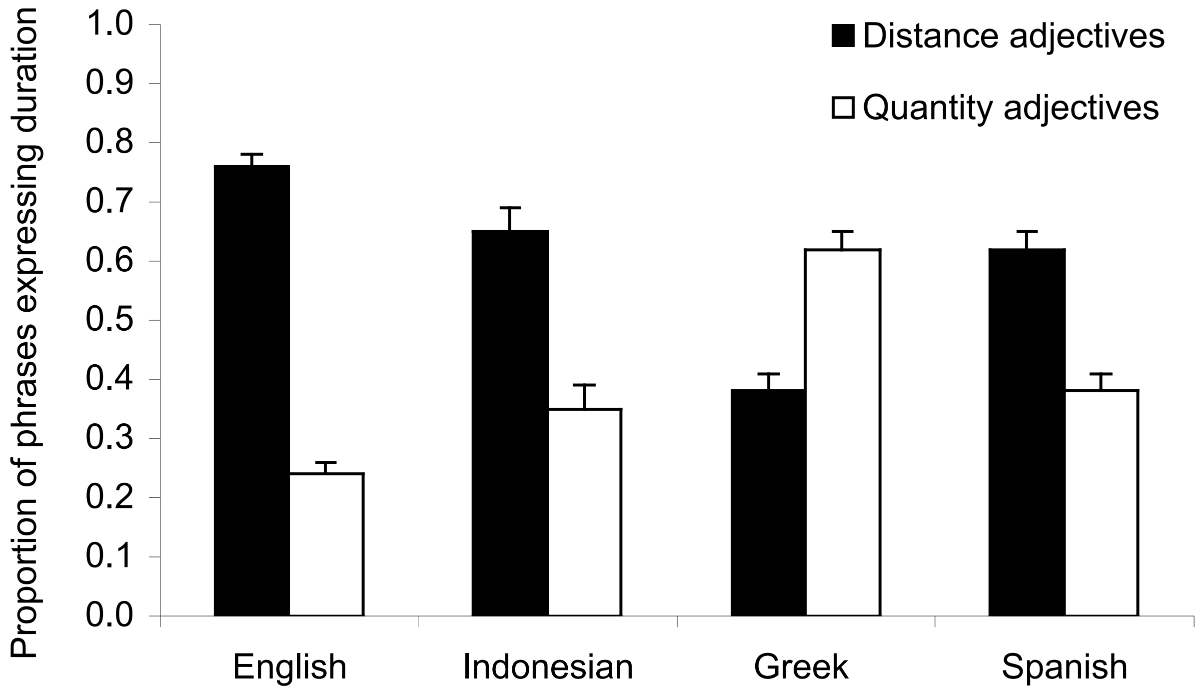

Figure 26. Results of questionnaire comparing the use of Distance and Quantity adjectives to modify event duration across languages.

Appendix 2

Upward Growing Lines in speakers of Indonesian, Greek, and Spanish

Instructions for Experiment 7c (Upward Growing Lines) were translated into Indonesian, Greek, and Spanish, and the task was administered to native speakers of these languages. Results for Indonesian, Greek, and Spanish speakers are summarized in figures 27a-d, 28a-d, and 29a-d, respectively. Results from English speakers are summarized in figure 10.

The slope of the effect of distance on time estimation was compared across languages (figure 30) using a one-way ANOVA ($F(3,46) = 1.22$, $p=.31$). Although not mandated by the nonsignificant ANOVA result, pairwise independent sample t-tests were conducted to confirm that slopes did not differ significantly between language groups.

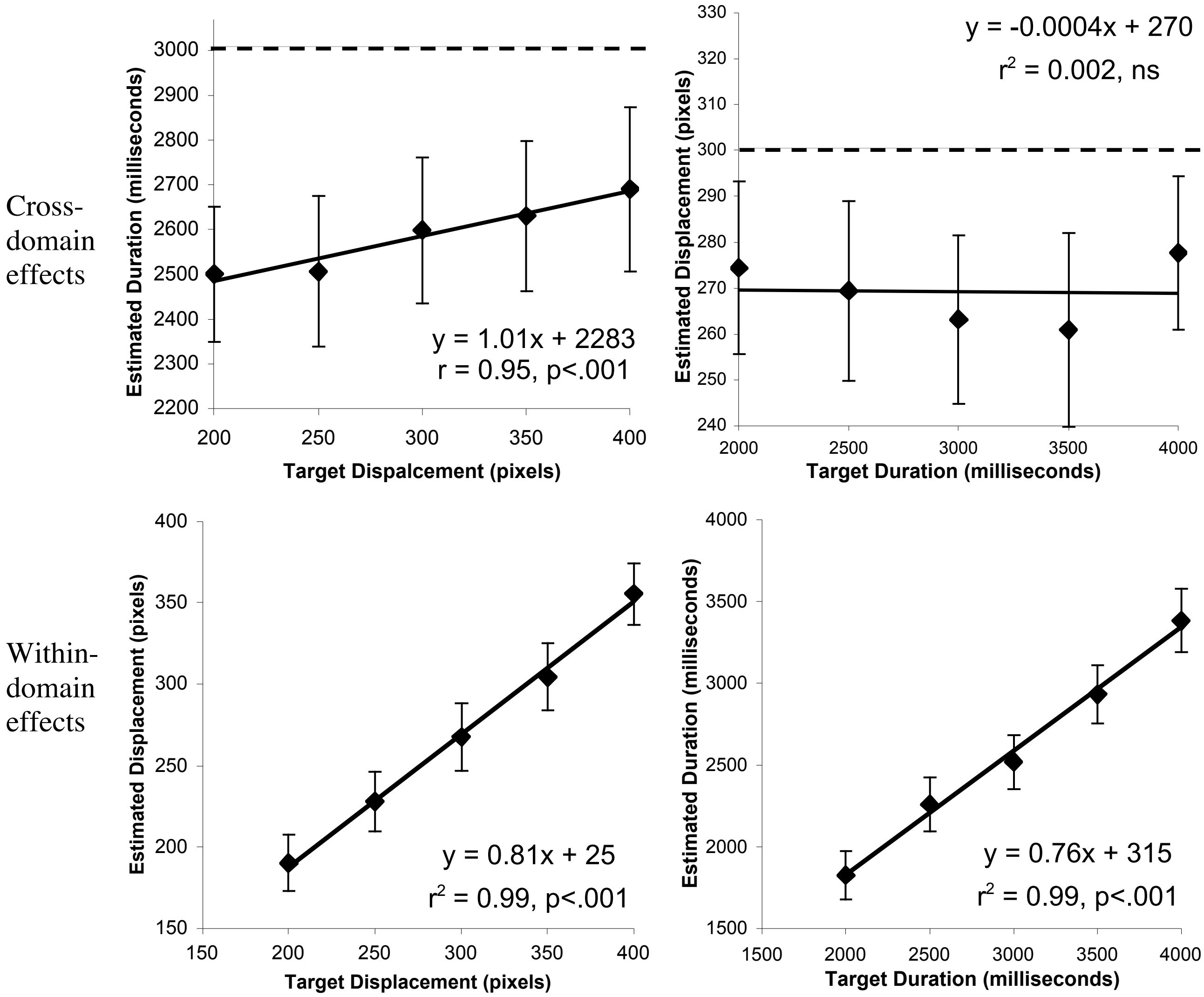

Figure 27. Grand averaged duration and displacement estimates (n=11) for the Upward Growing Lines experiment in Indonesian speakers. Top: Cross-domain effects. 27a. (left) Effect of displacement on duration estimation. 27b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 27a. and 27b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 27c. (left) Effect of target displacement on estimated displacement. 27d. (right) Effect of target duration on estimated duration.

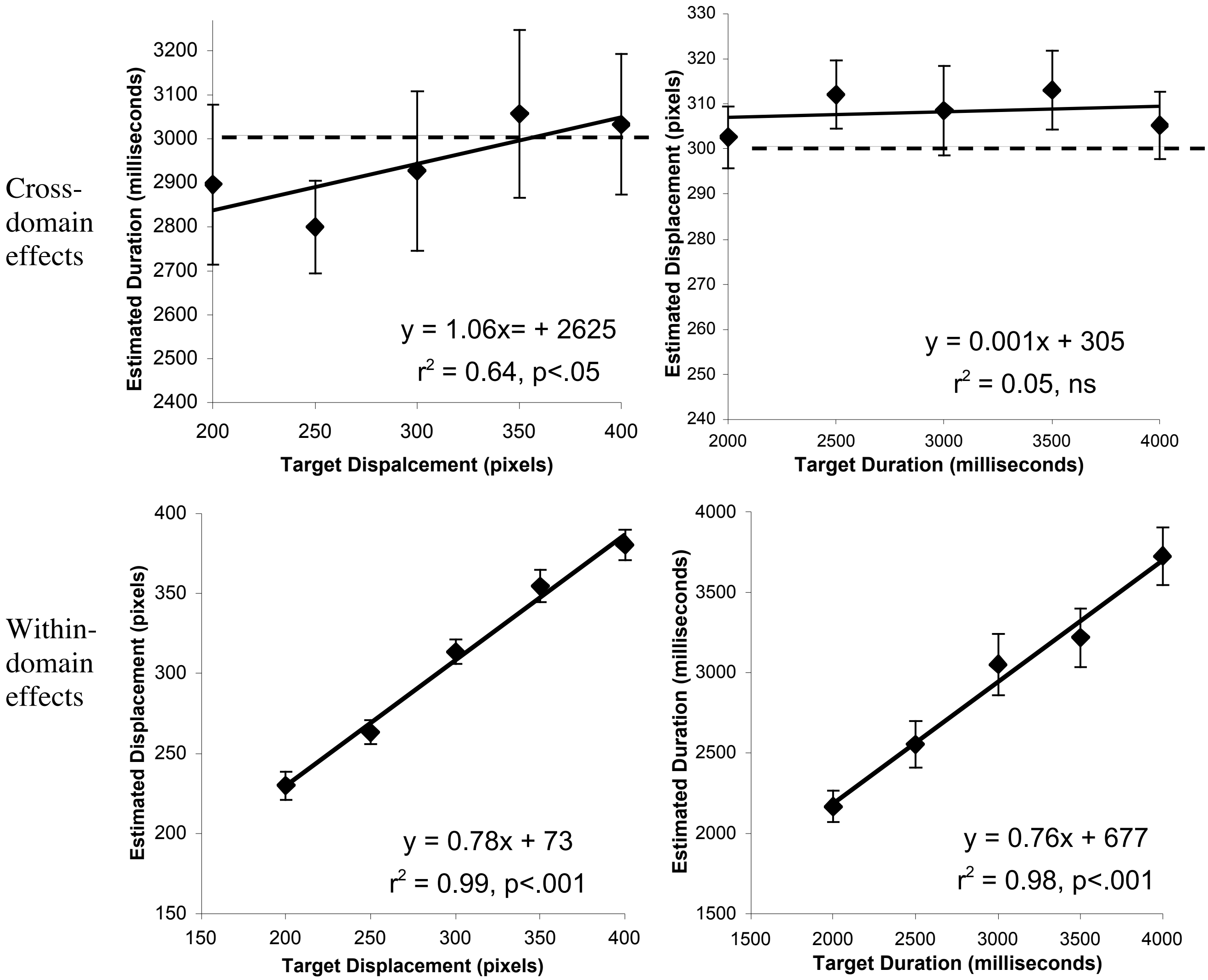

Figure 28. Grand averaged duration and displacement estimates (n=15) for the Upward Growing Lines experiment in Greek speakers. Top: Cross-domain effects. 28a. (left) Effect of displacement on duration estimation. 28b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 28a. and 28b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 28c. (left) Effect of target displacement on estimated displacement. 28d. (right) Effect of target duration on estimated duration.

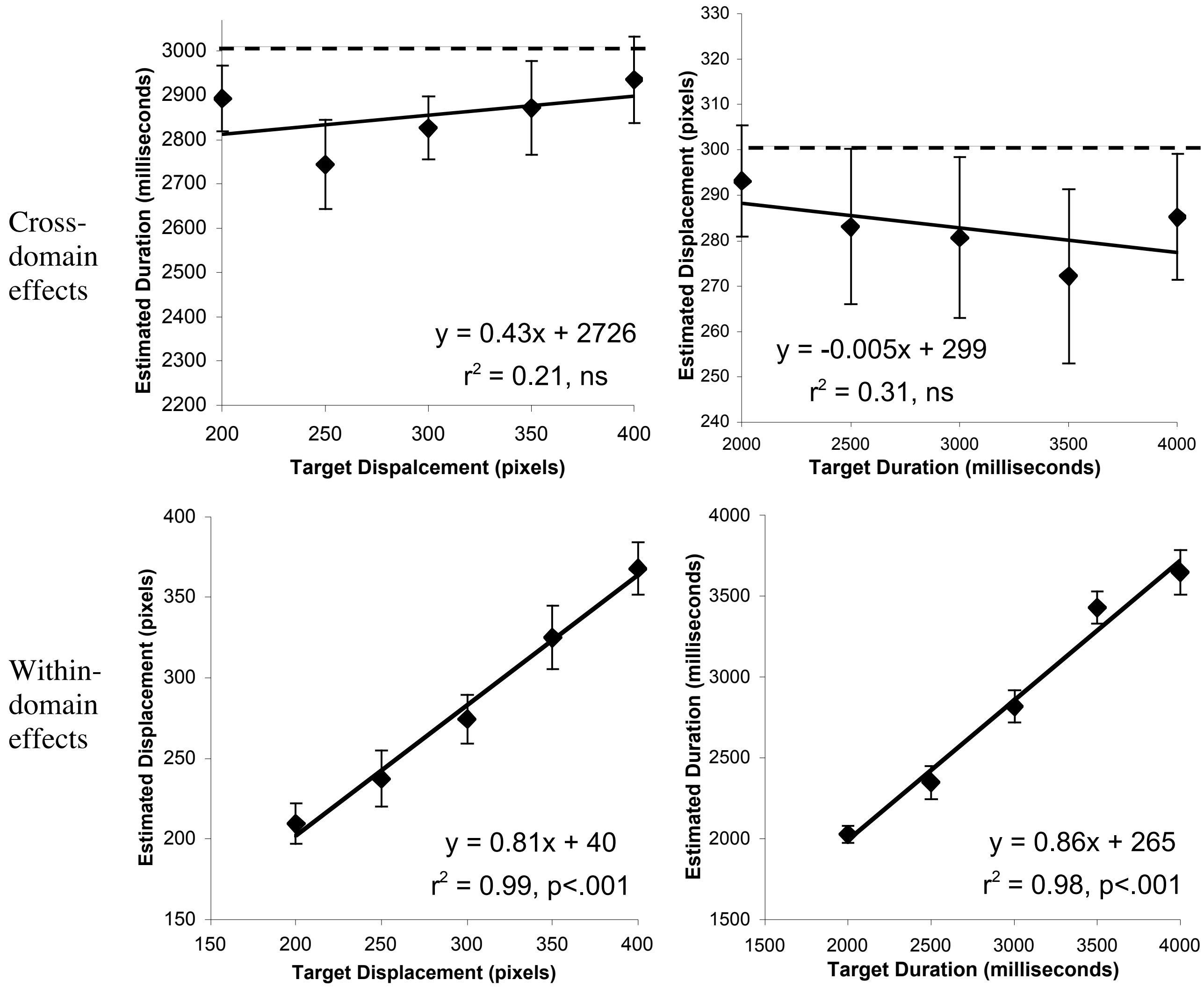

Figure 29. Grand averaged duration and displacement estimates (n=10) for the Upward Growing Lines experiment in Spanish speakers. Top: Cross-domain effects. 29a. (left) Effect of displacement on duration estimation. 29b. (right) Effect of duration on displacement estimation. The horizontal dotted lines indicate perfect performance. The ranges of the ordinates of 29a. and 29b. are proportionate with respect to the total range of target durations and displacements. Bottom: Within-domain effects. 29c. (left) Effect of target displacement on estimated displacement. 29d. (right) Effect of target duration on estimated duration.

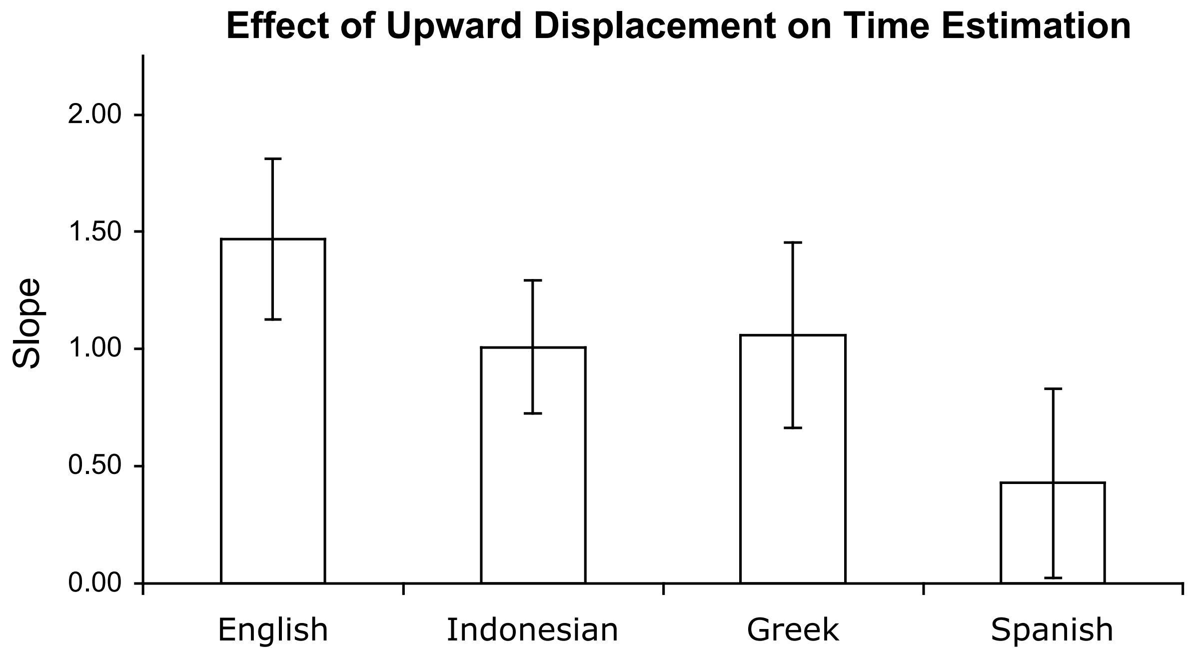

Figure 30. Summary of the effects of distance interference on duration estimation for the Upward Growing Line experiment (Experiment 7c) in speakers of English, Indonesian, Greek, and Spanish. The slope of the effect of target displacement on estimated duration did not differ significantly across language groups.